Texture classification and segmentation algorithms in man and machines

TERRY CAELLI
Department of Computer Science, The University of Melbourne, Parkville, Victoria 3052, Australia

Received for publication 7 September 1993

Abstract—In this paper an attempt is made to review the types of processes developed by the author, and others, to classify and segment textures and to compare, where possible, such algorithms with what has been proposed to occur in human vision. In particular, this paper is concerned with evaluating the proposed constituent processes from a 'cognitive engineering' perspective where, in order to have an adequate model for either human or machine texture processing, complete segmentation or classification or both must be attained.

INTRODUCTION—WHY STUDY TEXTURES?

When we, or machines, visually process what is sensed from the environment we perceive structures such as other flora and fauna, geophysical and man-made objects—all of which are depicted by regions in the sensed image. Most of these regions can be isolated (segmented) and even identified (classified) by the specific existence of differences in colours (individual position (pixel) intensity values) or surface textural characteristics. In the former case, 'first-order' region descriptors, such as the means, variances of red—green—blue (or appropriate colour channel) pixel values, probably account for most of the important region types. This follows from the fact that many object parts or objects, per se, are differentiated by their surface albedo and surface shape (depth modulations) and that, with illumination, the resultant images are consistent with these different surface properties.

However, there are surfaces which cannot be segmented or identified by these individual pixel values. Naturally occurring examples include tree bark, leaf striations and animal skins. With such cases it is the spatial pattern of light variability—texture—which defines the image region corresponding to the surface, whether it originates from the surface colour pattern or from the regular variations in the surface depth which, under illumination, produces textured images. Since these characteristics involve determining the relationships between intensities at different positions they are termed 'higher-order' and, where pair-wise relations are the basic unit of comparison, they are 'second-order'. These types of image features—texture features—are difficult to conceptualise and experimentally study, and over the past 20 years or so, there have been many attempts to understand how humans process such information to classify (identify) or segment (discriminate) regions in scenes.

Texture processing is also representative of the evolution of theory in biological and machine vision. This is because, first, texture processing covers core problems in the development of robust image segmentation algorithms and is basic to the
'early' or 'pre-attentive' figure/ground distinction in human vision. Second, the types of discussions and treatments of encoding and procedural aspects of texture processing are representative of a larger literature in 'early' vision. Consequently, the types of issues to be discussed in this article, it is hoped, will bear on more issues than the evaluation of encoding and algorithmic aspects of texture classification or discrimination, per se.

Indeed, my thesis has been for a while now (see Caelli and Oguztoreli, 1988) that human visual information processing has the following characteristics. It involves adaptive encoding strategies, where the features employed to solve one problem, for example, are not the same as those used in solving others. It includes considerable additional processes to those of feature extraction, and it typically involves prior knowledge and feedback from performance to lower level processes. In this paper I will endeavour to show just how these principles apply to texture processing and segmentation from work which has been completed by us over the past 15 years. This is not to imply that our work is unique. Rather, it is just representative of results in the area and references to related work is given where possible. But first, some history.

BRIEF HISTORY

Up to the late 1970s most approaches to understanding texture processing were focused on 'global' encoding models. That is, texture classification was determined from the similarities or differences between texture features necessarily computed over texture regions. Texture encoders typically involved both first- and second-order region statistics encapsulated by region intensity histograms, power spectra, autocorrelation or dipole measures, co-occurrences of pixel values. An example of the type of experiment that emerged from this approach would be the predicted similarity of two textures in terms of the similarity of the Fourier power spectra of each texture—the Fourier power spectrum being a global texture feature vector (whose components are the amplitude values at each frequency).

During this period of development, the 'texture paradigms' were of two basic types. One, the 'proof by contradiction' approach, particularly pioneered by Julesz (1962, 1982), and encapsulated by the notion that those texture characteristics, and their associated measures, which were relevant to texture discrimination, satisfied the rule:

If the feature's states are different and the textures are not discriminable, then the feature is not encoded.

Many examples of this type of conjecture were provided from the use of second-order statistics. The second paradigm, more often used in machine vision, was that based upon the traditional pattern classification rule:

Discrimination is monotonic with feature measure differences—so long as they are relevant to the information processing task.

These two paradigms motivated many experiments and investigations into texture discrimination and classification and, to this day, the latter paradigm still applies—being a generalization of the former.

The main problems with these earlier global representations were that the features, having been derived through integration over image regions, typically could not uniquely define the textures and so, inevitably, counterexamples could be found