Understanding vision – just around the corner or a distant dream?

Andrew Glennerster
School of Psychology and Clinical Language Sciences
University of Reading

Understanding vision is one of the central goals of neuroscience. Over the past two or three decades, there has been an explosion of research into the mechanisms of human and primate vision. Computational modelling has been of critical importance in making sense of the data, allowing researchers to draw generalisable conclusions and relate experimental findings to wider principles of information processing. Li Zhaoping's new book, *Understanding Vision*, offers a detailed and comprehensive introduction to this computational approach to vision research. It will be a valuable guide to anyone who wants to learn about the cortex as mechanism designed to process information efficiently. It provides a wealth of examples from human and animal experiments to illustrate the computational principles. But this book is not for the mathematically faint-hearted. To appreciate it, you will need to take in at least some of the equations. Nowadays, a remarkable number of researchers in neuroscience have a background in physics, maths or engineering and, for them, this book is an ideal bridge to the world of biological information processing.

The central claim is that vision is comprised of three stages: encoding, selection and decoding. Zhaoping illustrates all of these using examples from the primary visual cortex, V1, which is the first area of cortex that receives visual input and also the area that Zhaoping has studied most intensively in her research. Not everyone would agree that vision is best described as encoding, selection and decoding and many would choose not to emphasise V1 as much as Zhaoping does. Nevertheless, this book is a *tour de force* of much of the computational literature on visual processing. It will prove to be a very valuable resource for anyone with a numerate background who wants to learn about sensory neuroscience and anyone who appreciates a consistent mathematical framework applied to a wide range of topics in cortical visual processing.

Of the three stages - encoding, selection and decoding – the first is least contentious. Here, Zhaoping does an excellent job of bringing out the computational principles that underlie efficient coding, demonstrating how these are achieved in the early stages of visual processing. The figures are clear and well-crafted; they help the reader to get an intuitive understanding of the processing defined in the equations. For example, a figure might consist of several equations in the first column, an illustration of each in the second column, using a scatter plot to show how variables in the equations relate to one another, and a final column of images showing the effect of the operations defined by each equation. There is an impressive coverage of topics in this first section, including efficient sampling, efficient encoding, reduction of redundancy, adaptation to the statistical properties of the input and multi-scale image analysis. These are general problems of image encoding and they apply as much in machine vision as in biology. One issue about encoding that is raised but not fully answered concerns the relative *inefficiency* of V1 processing compared to the extraordinarily efficient spatio-temporal coding of signals in the retina. The tight informational bottleneck of the optic nerve, which has only 1 million fibres and could not increase in size without restricting the movement of the eyeball, imposes a severe practical constraint on efficient coding. On the other hand, in the cortex principles other than coding efficiency, at least in this strict sense, must take precedence.

The second stage, selection, is an area in which Zhaoping has worked extensively and this monograph-like chapter takes up more than a third of the book. It lays out her hypothesis that V1 provides a bottom-up 'saliency map' whose function is to alert higher areas to parts of the image that merit further attention. There are a series of psychophysical experiments on the detectability of targets in fields of distractors and a model of the connections in V1 that would explain why some targets and not others are easy to spot. A possible criticism is that there is only a limited discussion of rival models of bottom-up saliency and the chapter is very focused on V1. Maps encoding the salience or behavioural significance of visual targets exist in many other areas including pulvinar, superior colliculus, frontal-eye fields and lateral intraparietal area (LIP) and a broader discussion of visual selection, including processing in these areas, might have been helpful. Nevertheless, Zhaoping makes a clear and thorough case for one particular model.

The third stage, decoding, is more problematic. Zhaoping takes the reader through standard techniques for choosing between pre-defined interpretations of an image, such as Bayesian inference and maximum likelihood discrimination. She also emphasises that the decoding should be task-specific. For example, confronted by the same scene, different image parameters may be more or less relevant in different circumstances: motion flow is likely to be critical if the person is being chased by a dog; subtle colour distinctions are more important if they are picking fruit. The difficulty comes when Zhaoping attempts to tie these processes down to particular parts of the cortex. While the algorithms discussed are detailed and pertinent, the proposed neural mechanisms are sketchy at best. The book ends by suggesting that additional layers in a network (simple cells, complex cells, 'object units', etc) can carry out the operations necessary for decoding but this leaves the distinction between encoding and decoding rather ambiguous, at least at a neural level. For example, Zhaoping discusses the decoding of photoreceptor responses to light of different wavelengths. It is not clear how the next layer of V1 might carry out this decoding and how the neural operations would be different from the

encoding that V1 is also engaged in. Picking out relevant parameters from an image, or set of images, to control the current action or determine the next one is clearly a critical aspect (possibly *the* critical aspect) of visually guided behaviour but it seems unlikely that stacking layer upon layer of retinotopically organised networks will perform this function. At any rate, the book does not explain how this could work.

Understanding vision is hard. We walk around, moving our head and eyes freely and yet our perception of the world remains stable; effortlessly, we pick out the relevant visual information we need to guide our actions. None of that can be understood by considering just the output of V1 or, indeed, of any retinotopically organised visual area since these outputs change dramatically every time we make a saccade. Understanding vision requires a thorough grasp of the tasks that animals are engaged in and the type of representations they store. Many sensory modalities must contribute to these representations and they must enable the animal to move smoothly from one component of a complex sequence of actions to the next. However much we discover about the details of V1 and similar visual areas, it is unlikely to help us find a solution to these puzzling problems.