Chapter 1
Understanding vision from images to shapes

1.1 Introduction

What does it mean, to see? Most people would, I think, agree that the act of seeing is the process of discovering from the images falling on our retinas what is actually present in the world, and where it is. This entitles us to regard vision first and foremost as an information processing task, but we cannot think of it just as a process. For if we are capable of knowing what is where in the world, our brains must somehow be capable of obtaining not only this information from images, but also of representing it—in all its profusion of color and form, beauty, motion and detail. The study of vision must therefore include not only the study of how to extract from images the various aspects of the world that are useful to us, but it must also inquire into the nature of the internal representations by which we capture this information, and thus make it available as a basis for decisions about our thoughts and actions.
From this rather philosophical point of view, therefore, understanding how we see consists essentially of understanding how our brains obtain and represent information about the world from images of it. But what precisely does understanding mean here? There is a whole range of perspectives that must be satisfied before one can be said, from a human and scientific point of view, to have understood visual perception. First, and I think foremost, there is the plain man. He knows what it is like to see, and unless the bones of one's arguments and theories roughly correspond to what he knows to be true at first hand, one will probably be wrong (a point made with force and elegance by the philosopher L. Austin, 1962). Secondly, there are the brain scientists—the physiologists and anatomists—who know a great deal about how the nervous system is built and how parts of it behave. The issues that concern them—how the cells are connected and why they respond as they do—must be resolved and placed in their proper perspective in any full account of perception. And the same is true for the experimental psychologists.

On the other hand, someone who has bought and played with a small home computer may make quite a different argument. If, he might say, vision really is an information processing task, then I should be able to make my computer do it, provided that it has sufficient power, memory, and some way of being connected to a home television camera. The explanation he wants, therefore, is a rather abstract one, telling him what to program, and if possible, a hint about the best algorithms for doing it. He doesn't want to know about rhodopsin, or the lateral geniculate nucleus, or inhibitory interneurons. He wants to know how to program vision.

This brings us to a fundamental point; namely, that in order to understand fully a device that performs an information processing task, one needs explanations. T. Poggio and I formulated this in terms of three levels of explanation. At the first level is the computational theory of the process. Here is formulated the goal of the computation explaining why it is appropriate and setting out as rigorously as possible the logic of the strategy by which it may be carried out. For example, this level of explanation applied to the checkout computer at a supermarket would involve showing it appropriate that the computer should perform addition (instead of mul-

---