

## *Chapter 1*

# Introduction

*I stand at the window and see a house, trees, sky.  
Theoretically I might say there were 327  
brightnesses and nuances of colour. Do I have  
"327"? No. I have sky, house, and trees.*

*Max Wertheimer, 1923*

The traditional goal of computer vision was to deduce the properties and structure of the three-dimensional world from one or more two-dimensional images, constraining the field to problems in geometry (e.g. Marr, 1982). As our understanding of the biological visual system has grown, computer vision theorists developed a new goal: to automate the visual recognition of objects and the understanding of complex scenes. Such research can not succeed within the field of computer science alone. It must also draw on research efforts in neurophysiology, neuropsychology, psychophysics and psychology.

Some success has come from the study of “low-level” vision across these fields. For example, it is generally accepted that primary visual cortex (V1), the first cortical

brain area involved in processing the visual stimulus, is made up of neurons tuned for different spatial frequencies and orientations across the visual field. This conclusion is supported by physiological mapping of the receptive fields (RFs) of these neurons (e.g. Hubel & Wiesel, 1979), and a multitude of psychophysical tests (e.g. Campbell & Robson, 1968). The computational theory might then be that V1 produces an edge map of the viewed image. Edges contain rich information about the shape and location of objects in an image, so the representation would be useful for later visual areas.

The aim of this thesis is to initiate investigations of “higher-level” aspects of vision such as the categorization and recognition of objects and scenes. As we move “higher” in the visual stream, to later visual areas and more complex processing, decoding the neural signals becomes more challenging. Psychophysics gets more difficult as the behavior of subjects grows more complex. Computer vision becomes an inventive enterprise in the absence of underlying theories of visual processing. By combining efforts from these two fields in a single research project we may make faster progress towards understanding visual perception. High-level behavioral data can serve as a “benchmark” against which the performance of algorithms can be measured. The details of the algorithms can inspire questions about visual routines that should be tested psychophysically.

As noted by Egon Brunswick (1947), a pioneer of ecological approaches to perception, what we learn about the visual system in a limited domain can only be applied to that domain, and can not be generalized directly to other domains. If we wish to learn how the visual system processes complex natural stimuli for the purpose of visual search, motor interaction, social recognition, etc., it is necessary to make the stimuli and

the tasks in our experiments as realistic and natural as possible. This general guideline has motivated the design of the following research.

Object recognition and scene understanding both occur later during the time-course of visual processing and they require the bottom-up input of image features as well as the top-down input of memory representations and expectations. We still lack a theory for how objects and scenes are encoded, but it has been suggested in neuropsychology that the spatial relations between the parts of a single object are encoded independently and in parallel to spatial relations between perceptual objects in a scene (Humphreys & Riddoch, 1995). Neural imaging results support this idea by proposing separate cortical areas for object and scene identification (Epstein & Kanwisher, 1998).

The flow of this thesis is built around the hierarchical, scale structure of the perceptual image: scenes are made up of objects and objects are made up of parts. In addition to their part structure, objects and scenes possess holistic properties that the visual system can use. Chapter 2 will investigate the rules that govern our perception of object parts. Chapter 3 makes use of eye movement studies to explore that nature of our object memory representations. Chapter 4 asks whether or not mathematical formulations of shape, a holistic object property, can be used to measure perceptual shape similarity. Chapter 5 proposes a holistic cue that may be useful for rapid scene identification.