1. Computational Vision in Neural and Machine Systems

Michael Jenkin and Laurence Harris

1.1 Introduction

The ability to process visual information streams is a critical requirement for both biological systems as well as for a wide variety of robotic devices. The fundamental need for effective visual information processing in biological systems is illustrated in Figure 1.1. These three snapshots show a cheetah emerging from the long grass in Tanzania. Being able to discern the shape emerging from the tall grass early is a critical survival skill. Biological systems that are unable to process the visual information in a timely fashion are unlikely to succeed in the wild. Similarly, in the machine vision domain, timely effective processing of visual information is often key. Figure 1.2 shows the AQUA robot, a visually-guided amphibious robot that is capable of unsupervised operation (Dudek et al., 2005). This vehicle relies primarily on visual information obtained from forward facing cameras to reason about its external environment. Without the ability to process its multiple video camera inputs in a timely fashion, the robot would be unable to operate.

In 1991 (over 15 years ago), The Centre for Vision Research at York University held the first in what has become a bi-annual conference on vision. The 1991 conference – which resulted in the book *Spatial Vision in Humans and Robots* – examined how biological and machine systems address the task of processing the rich visual field in order to recover information about the spatial surround. Fifteen years later, the York Vision Conference has re-visited this fundamental issue: the relationship between computational models of visual information processing and research into biological visual information processing.

The last fifteen years has seen astounding advances in both artificial and biological vision. Driven (at least in part) by advances in the technology available to explore how biological systems process visual information, and the orders of magnitude performance improvements in the computational power that can be brought to the task of processing visual imagery, visual models have advanced from processing a single image viewed in isolation to a stream of embedded visual information processing within
Figure 1.1: Vision is a critical perceptual ability for many biological systems. Early detection of visual events – such as that depicted above – can be essential for an individual’s survival.

an ongoing spatio-temporal relationship. In the computational field, this has lead to the consideration of visual information processing not as the evaluation of an isolated static image, but rather as the task of processing an image stream within the context of some wider task. In the biological fields, this has lead to a wide range of advances including the emergence of models of multi-modal fusion of information from different perceptual systems.

As visual information processing is considered within a temporal context, many of the problems that occur “naturally” in the biological community become apparent in the computational one. This includes tasks such as integrating information from multiple
views, searching for specific objects within a wide visual display, and attending to salient features within the environment.

David Marr (1982) distinguished three levels of visual processing: computational, algorithmic, and implementational. His computational level consisted of a description of what computation needs to be performed and what information is available to perform the computations on. His algorithmic level specified how the computational level might be performed. Algorithms performed biologically are likely to be very different from those performed on a computer. For example there are many levels of parallel processing in the brain, which is unusual in a digital computer. The implementation level is the act of performing the selected algorithm, either in the brain or in a digital computer. This book places more emphasis on the first of Marr’s three stages, outlining the principles of the computational processes to be performed with less emphasis on the actual algorithms that might be employed to run them.

This volume is divided into three parts, centred around the topics of dynamical systems; attention, motion and eye movements; and stereo vision. Dynamical systems deals with adaptation, motion detection, robotic vision systems, shape recovery from image sequences, and the reconstruction of objects from parts and attributes processed separately. The section on attention deals with attention and action, visual search in clutter, the memory of visual features across saccades, and modelling gaze in natural images. Finally, the section on stereo describes a number of algorithms and approaches that reflect the current state of the art in stereo vision algorithms and models of stereo information processing.

In each of these sections we find papers that examine spatial information processing and how it interacts with the temporal domain. In Part I, for example, Norma Graham and Sabina Wolfson examine specific adaptation processes in human visual information processing. They question how various levels of visual information processing adapt to the absolute levels of illumination that are available and the time course of this adaptation. In Part II, Steven Prime, Matthias Niemeier and Douglas Crawford examine how visual information is maintained across visual saccades, a problem that is critical to biological systems that utilize eye movements to integrate larger portions of the visual field than are available in a single gaze, and which is also critical in machine systems which must use camera and vehicle motion to deal with the limited field of view of existing camera technologies. Finally in Part III, Jane Mulligan examines how stereo image processing can be made sufficiently “computationally efficient” that it can be embedded within machine vision systems and used as a building block for telepresence systems.

Beyond these three examples we find a collection of chapters that seek to address spatial vision information processing in both the computational and biological fields. These chapters illustrate just how detailed our understanding of basic visual information processing has come, and how much remains to be discovered. They also demonstrate how similar the problems are that are encountered by biological and computational systems and how similar are the underlying information processing models (algorithms). Much has been accomplished in the 15 years since the first York Vision Conference on Spatial Vision in Humans and Robots. As we observed in the introduction to the book that arose from that conference, the two communities can learn a great deal from each other. That observation seems just as true today.
1.2 The CD-ROM

Enclosed with this volume is a CD-ROM that contains video, colour imagery, and other digital media associated with the text. A complete copy of this volume in PDF format can also be found on the CD-ROM. The material on the CD-ROM can be accessed using a standard browser (such as Internet Explorer or Firefox). Videos on the CD-ROM are viewable with Quicktime, while viewing of the presentations on the CD-ROM will require a PowerPoint viewer.

References

