Processing Motion in the Real World

Johannes M. Zanker and Jochen Zeil

Visual Sciences Group, Research School of Biological Sciences, Australian
National University, Canberra, ACT, Australia

1. Observing motion

As Nick Wade illustrates in wonderful detail in his book on the history of visual science (1998), it was the development of experimental sciences that converted scholars from outdoor observers into laboratory experimenters. This novel approach opened new opportunities to study nature, and since the 17th century advanced knowledge in visual sciences as in so many other areas with impressive speed. Experimental sciences, combined with the theory of evolution, opened a new perspective by comparing design principles for similar sensory and behavioural functions across different species. Carrying out critical, theory-driven experiments under highly controlled conditions is so powerful, compared to pure observation of nature, that euphoria about precise and clean evidence made scientists sometimes forget the conditions under which such evidence is collected. The visual stimuli used in laboratory experiments are by necessity rather “abstract”, simple, and barely representative of the natural operating conditions under which biological systems have evolved and continue to operate. The focus on laboratory studies has dominated visual science ever since, although in the work of some scientists observation and experimentation had always been part of a unified approach. All the way along there have been continuing attempts for instance, to understand of human perception in its real life context (a number of “classical” examples can be found in Helmholtz 1864; Metzger 1975; Gibson 1979).

This book tries to pull together research from very different areas of visual sciences with the aim of evaluating whether the developments in knowledge and technology during the last decades allow us to consider visual motion processing under conditions that have been relevant in evolution. What is the framework for such a paradigm shift from laboratory-based investigations of visual systems to their analysis in the real world?

- *Technological advances.* Recording devices with high resolution in time and space allow us now to monitor behavioural patterns in the field with hitherto unknown precision. The availability of digital cameras and powerful storage devices at a reasonable price furthermore opens the possibility to record the structure and dynamics of natural habitats at a large scale, with the option to analyse the spatial distribution of light together with its spectral composition and its polarization.
- *Theoretical approaches.* Motion detector theory and simulation devices have been developed to a stage at which responses to complex inputs can be analysed. At the same time, theoretical approaches to understanding neuronal coding strategies have reached a solidity and sophistication that encourage us to move away from interpreting average responses to average stimulus conditions. Understanding responses to a single stimulus presentation may offer the chance to trace a behavioural decision in a visuo-motor task down to the level of the individual neurones in a complex neuronal circuit.
- *Neurophysiology and Behaviour.* The extensive knowledge of the computational properties of individual neurones and the interactions in neuronal circuits in the visual pathways provides us with a solid basis to ask how the image processing of natural scenes is limited by the structure and function of real brains. There are recent attempts to record nerve cell activity under more and more natural conditions, leading eventually to field electrophysiology. One of the goals that may be achieved in the nearby future is to study the intricate relation between motor and sensory patterns emerging during active vision in its natural context.
- *Robotics.* Advances in systems control theory, opto-electronics and micro-mechanics have led to rapid developments in robotics so that we can now test neuroethological concepts in the closed sensory-motor loop. At the same time neurobiological knowledge can be used to develop robotics, as demonstrated by the emerging discipline of “biomimetics”. Implementations of the principles of visual information processing that have been identified in biological systems can be put to test in artificial systems in a variety of platforms and environments.

In this book, we try to illustrate how far we have come in our understanding of motion vision, and where the essential advances could be expected in the future. Motion vision is special because it is a non-trivial visual processing task and has a high “information content” for any organism (Borst and Egelhaaf 1989). Motion information is used to control orientation and movement, to relate to other animals of the same or other species, and to extract information on the three-dimensional structure of the environment (Nakayama 1985). It is thus not surprising that motion vision is highly developed in basically all diurnal creatures. Consequently, motion vision has been extensively studied in neuroethology (e.g. Miles and Wallman 1993) and psychophysics (e.g. Ullman 1979; Braddick and Sleigh 1983; Smith and Snowden 1994; Watanabe 1998), and continues to be a challenge for

machine vision (Marr 1982; Aloimonos 1993; Srinivasan and Venkatesh 1997). Recognizing the recent achievements of multidisciplinary research on motion vision, this book brings together authors from various disciplines ranging from engineering and biology to psychology.

2. Processing constraints

We begin our survey with the fundamental neuronal mechanisms of motion detection and the integration of local motion information. On this basis we discuss a number of essential control problems that are solved by using motion information, and ask how the motion signals necessary for such control systems can be encoded with neuronal processing elements that have limited bandwidth and reliability. Three types of constraints, which limit the performance of biological signal processing systems, are considered in this book.

- *Computational*: The visual system has to extract relevant information about egomotion, about the three-dimensional layout of the environment, and about moving objects, from complex, dynamic, two-dimensional images (Gibson 1979). In each case it has to cope with highly ambiguous data. Well known examples of such ambiguities are the so-called “aperture” or “correspondence” problems which arise at the elementary level of motion detection (Marr 1982; Hildreth and Koch 1987). They demonstrate that the basic computational problems in motion vision are mathematically ill-posed.
- *Neural*: Biological systems perform computations with neurones which suffer from a number of severe processing limitations. Neurones possess a comparatively small dynamic range for representing intensities and temporal changes with analogue neuronal signals or spike trains (Barlow 1981), they can only approximate exact mathematical operations (Torre and Poggio 1978), and suffer from internal noise (Bialek and Rieke 1992; Laughlin 1998).
- *Ecological*: Visual systems operate in concrete and often very specific worlds, which are characterized by differences in the structure of behaviour and the topography of the environment (Lythgoe 1979; Dusenbery 1992). Given that motion processing mechanisms have evolved under selective pressure in specific visual habitats and in the context of specific lifestyles (O’Carroll et al. 1996), the systematic analysis of visual environments and visual tasks should help us to understand the functional and adaptive properties of neural processing strategies. At the moment we have surprisingly little to say regarding the question that arises repeatedly in this book: what are the actual motion signals visual systems have to work with?

3. Steps to analyse a complex system

The book deals with six major topics to investigate the significance of these constraints for motion vision. Each part is organized in a keynote chapter that introduces a topic and the crucial concepts, which are expanded, complemented or juxtaposed by shorter companion articles that provide additional or alternative views on the same topics. This format naturally does not cover the field comprehensively, but hopefully offers the reader a multi-faceted insight into a set of questions that need to be addressed when we try to assess our knowledge of visual motion processing in natural environments.

(I) The first part of the book deals with the biological basis of motion detection from a physiological and anatomical point of view. It lays out what we know about how motion detection is implemented with neurones and synapses, and how their properties can be related to theoretical models of elementary motion detection, which is fundamental to all consecutive processing. How do neurones perform the basic mathematical operations that are necessary to extract directional selective signals from the spatial and temporal changes of image intensity (Reichardt 1987)? The review by Vaney et al. describes in great detail the neuronal machinery of such a spatiotemporal correlation mechanism for the rabbit retina, which now has been studied over decades. Although such cellular models of connectivity are now described down to the level of the biochemistry of the synapses, it is surprising to see that some essential questions are still a matter of debate. The neuroanatomical and functional structure of directionally selective ganglion cells in the rabbit is compared in the companion chapters with two very different biological systems. Ibbotson illustrates how specific models of motion detection can be discriminated by careful experimentation in the marsupial, a comparatively distant relative of the rabbit, in which – like in many higher mammals, including primates – the elementary steps of motion detection are not carried out in the retina, but in the cortex. Despite a completely different localization within the visual processing stream, which involves different classes of neurones, the computational structure of the local motion detecting process is strikingly similar in the rabbit and the wallaby, apart from some specific variations in synaptic connectivity. The theme of functionally equivalent processing by very different neuronal elements is further developed in the contribution by Douglass and Strausfeld who review the anatomical knowledge of motion detection networks in the visual system of flies. Neurones in animals with widely different phylogenetic history have such similar functional properties in the context of motion detection that they have even been assigned similar labels such as “magnocellular stream”. We have thus to appreciate that computational needs have recruited very different neuronal substrates in the course of evolution to solve one and the same task.

(II) Although we know the fundamental principles and the biological realization of the initial stages of motion processing quite well, we are far from understanding how behaviourally relevant information is extracted. The reason being that local

motion information, as it is extracted by elementary motion detectors, is noisy, ambiguous, or even misleading (Egelhaaf et al. 1989). The crucial information is often only carried by the whole distribution of motion signals (Koenderink 1986). The first processing stage in which such distributions can be extracted involves two fundamental classes of operations: image segmentation and spatiotemporal integration (Braddick 1997). The fact that we are able to perceive two motion signals that differ in direction or speed simultaneously within the same region of the visual field – a phenomenon called “transparency” – is often regarded as critical for understanding the competitive demands of integration and segmentation. Braddick and Qian discuss this topic from the viewpoints of both human psychophysics and primate electrophysiology. The authors address the question of how local motion signals are pooled across space and time while retaining sensitivity to different motion directions and indicate at which levels in the cortical processing stream the two mechanisms need to be localized. Braddick and Qian suggest that there must be an intermediate integration stage that has not yet been identified in terms of neurones. This role of motion opponency – being an essential part of local motion detection – for motion transparency is considered further by Zanker, who develops a computational model that accounts for specific properties of motion transparency and segmentation that are found in psychophysical studies. This model converges with the physiological considerations raised in the keynote paper, and predicts spatial constraints of separating motion signals. A more fundamental approach is taken by Cropper in his companion article to discuss the question how useful different kinds of motion stimuli are to study “global” motion percepts. He scrutinizes the variety of local features that can be used in a segmentation or pooling process, and asks to what extent our current experimental paradigms fail to address the question of how stimulus feature combinations are represented in cortical processing. These critical questions remind us that in trying to understand the interaction between motion signals we must be aware that brains are extremely powerful in combining information across modalities.

(III) More complex motion signal distributions, which have an extraordinary significance in everyday life, are the optic flow fields experienced by a moving observer. Extracting reliable information from optic flow is a crucial task for any mobile organism, because vision is required for the control of locomotion (Gibson 1979). The task of estimating egomotion parameters from optic flow (Koenderink 1986) is discussed in this part of the book from both biological and theoretical perspectives, in an attempt to characterize the operating principles. Dahmen et al. identify the principal limitations in comprehensive simulation experiments and then ask to what degree visual systems are optimized to extract egomotion parameters from optic flow. Simulations demonstrate how the fundamental algorithms can be realized by matched filters. By making assumptions about the statistical structure of the world and about typical patterns of locomotion, Dahmen et al. derive matched filters, which turn out to resemble the distribution of directional sensitivity of large field integrating neurones in the insect visual system. In

his companion paper, Perrone draws attention to the fact, that despite all their sophistication, current models continue to be hampered by the aperture problem that can lead to significant misjudgements of local motion direction. On the other hand, Srinivasan discusses how some invertebrates could use “quick and dirty short-cuts” to overcome some of the difficult problems of egomotion estimation. It is clear from the papers presented in this part that comparatively simple algorithms can be designed with pragmatic assumptions to analyse certain aspects of optic flow for a range of conditions, and that biological systems are experts in doing so. But the accuracy of the local motion information limits in various ways the precision that can be achieved, and under many conditions the visual system may need to work around typical pitfalls by using “rules-of-thumb”.

(IV) The next part takes a closer look at the intricate connection between the control of locomotion and the motion signals that are to be processed for this purpose. What are the fundamental computational strategies that are involved in gaze stabilization and tracking eye movements, and how is performance limited by the constraints imposed by the elementary motion detection process, and by the neural implementation? Sandini et al. focus in their keynote paper on the coordination of two eyes and the binocular integration of motion information that is needed to keep an object in the centre of the visual field. This task is complicated by geometrical aspects, such as translational components that result from excentric rotations, and particularities of motor dynamics. The use of mechanosensory cues can be very helpful in this context, and Sandini et al. demonstrate how both biological and artificial systems make use of such information. This relation between different sensory cues and the design of motor systems makes us aware of the need to consider the cross-modal context in which animals normally operate. One particularly interesting aspect is how a growing and learning organism adapts to the changes in perspective and the size and shape of its own sensory organs. The world seen by a newborn is not the world seen by a 20 year old and this again differs from people in their sixties. The visual field, for instance, has been reported to expand from a more ventral region to a more dorsal one, when human infants work their way up from a predominantly horizontal to an upright posture (Mohn and van Hof-van Duin 1991). We are far from understanding, however, the general patterns of perceptual adaptations that reflect changes in the visual environment. The chapter of Atkinson and Braddick addresses this issue by taking a developmental perspective of how the human visual system through growth, maturation and learning finally reaches the finely tuned visuo-motor coordination that we all rely on. Visual control of motor activity clearly has to be acquired when it comes to driving vehicles, and motion information is likely to play a crucial role in this task (Lee 1976). Land however provides a contrasting view to our mantra about the importance of motion information. He shows that during steering a car the control of gaze is not determined exclusively by the analysis of optic flow, but can be described in terms of simple geometric operating rules. Under natural operating conditions nervous systems thus exploit and combine sources of information that are useful and reliable.

(V) Up to this point we have treated motion information as if it was represented in some more or less instantaneous manner and independent of the immediate history of the sensory signals. Given the rapidly changing visual input in natural environments, this approach has two major limitations. It fails to take into account firstly, how adaptative properties of neurones change the instantaneously available information, and secondly with what precision the neural system can encode rapid changes. In biological systems, information is coded by neurones which have a limited bandwidth, in particular in the temporal domain. So what are the limits to the precision with which motion is represented by neurones? We had to discover in the preparation of this part of the book, that this question is by no means settled and therefore present highly controversial opinions – backed by elaborate experiments and sophisticated mathematical analysis – side by side. On the one hand, Warzecha and Egelhaaf argue in their contribution that in the motion sensitive neurones of the fly little or no information is carried by the exact timing of individual action potentials. Flies rely on temporal averages within a window of about 40 ms. On the other hand, de Ruyter van Steveninck et al. make a case that the same class of neurones exhibit extremely high precision in the timing of spikes. These opposing interpretations are derived from surprisingly similar experimental and theoretical approaches, in which, however, a crucial question remains unresolved, namely what exactly natural stimuli are. It is left to the reader to evaluate this pointed scientific discourse, in which cutting edge experimental and theoretical techniques are put to a biological reality check¹. A comparative component is added by the paper of Barberini et al., demonstrating that MT neurones employ similar coding strategies to those of fly visual interneurones, which takes us back to the theme of analogous implementations in different branches of the animal kingdom that has been touched upon in the first part. A final word of caution is added by Maddess who notes that adaptation of neuronal activity, on a variety of time scales, can change the information content of spike trains considerably. We thus have to realize that the attempt to assess “motion vision in the real world” needs to include the consideration of signal processing dynamics and coding limitations, which despite substantial theoretical and experimental advances in recent times are far from being understood.

(VI) Under natural operating conditions, an animal is not only exposed to the neural and computational constraints of motion vision, but also to those imposed by its habitat and lifestyle. Two aspects of natural operating conditions have to be considered: (i) the spatial and temporal distribution of biologically relevant signals in a given visual habitat and ethological context, and (ii) the structure of locomotion which to a large extent determines the pattern of motion signals an observer experiences. What do we need to know about lifestyle and the dynamic structure of the environment, and how can we relate such knowledge to

¹ The note added “in proof” (i.e., after the end of the refereeing process) by the authors of the keynote paper in response to the companion article illustrates the intensity of the continuing discussion of the issues covered by this part of the book

the neural and computational constraints? In their keynote paper, Eckert and Zeil make an attempt to compile a preliminary inventory of the relevant questions that need to be asked and of the available facts about the motion signals that animals experience under natural conditions. The paper emphasizes the fact that the major part of image motion is generated by animals themselves, so that the analysis of behaviour will play a crucial role in understanding the conditions under which motion processing normally operates. A challenging theme emerging from this analysis is that of characteristic motion habitats and their statistical properties: Even if environments are statistically self-similar, different animals have to attend to different relevant events, and – depending on their way of locomotion – will experience different “motion environments”. A more formal analysis of the spatiotemporal structure of image sequences is added by Dong in his companion chapter, who suggests characteristic coding strategies that resemble those proposed for achieving optimal representation of static images (Olshausen and Field 1996). Our last chapter provides neurophysiological evidence for motion processing being adapted to particular behavioural and ecological niches. By analysing large-field motion sensitive insect neurones, O’Carroll demonstrates how the structure of the sensory organs, the style of locomotion and the coding properties of neurones in insects reflect specific environments and lifestyles.

4. Conclusion

We are thus at a point where it becomes feasible, from a technical point of view, to describe and interpret biological visual systems in the context of their natural operating conditions. However, we have to acknowledge that we are just beginning to understand how neural, computational, and environmental constraints have driven the evolution of neuronal information processing mechanisms. It is only with a clear knowledge of these constraints that we can hope to develop smart machines which are as versatile, robust, competent, and flexible as the most humble animals evidently are.

References

- Aloimonos Y (1993) Active perception (Computer Vision). Erlbaum, Hillsdale
- Barlow HB (1981) Critical limiting factors in the design of the eye and visual cortex. *Proc Roy Soc Lond* 212: 1-34
- Bialek W, Rieke F (1992) Reliability and information transmission in spiking neurons. *Trends Neurosci* 15: 428-434
- Borst A, Egelhaaf M (1989) Principles of visual motion detection. *Trends Neurosci* 12: 297-306
- Braddick OJ (1997) Local and global representations of velocity: transparency, opponency and global direction perception. *Perception* 26: 995-1010
- Braddick OJ, Sleigh AC, eds (1983) Physical and biological processing of images. Springer Verlag, New York

- Dusenberry DB (1992) Sensory ecology. How organisms acquire and respond to information. Freeman and Co, New York
- Egelhaaf M, Borst A, Reichardt W (1989) Computational structure of a biological motion-detection system as revealed by local detector analysis in the fly's nervous system. *J Opt Soc Am A* 6: 1070-1087
- Gibson JJ (1979) The ecological approach to visual perception. Lawrence Erlbaum Assoc, Hillsdale, New Jersey
- Helmholtz HLF von (1864) Treatise on physiological optics. English translation from German (1962), Southall JPC (ed) Dover Publications, Dover
- Hildreth E-C, Koch C (1987) The analysis of visual motion: From computational theory to neuronal mechanisms. *Ann Rev Neurosci* 10: 477-533
- Koenderink JJ (1986) Optic flow. *Vision Res* 26: 161-180
- Laughlin SB (1998) Observing design with compound eyes. In: Weibel ER, Taylor CR, Bolis L (eds) Principles of animal design. Cambridge University Press, Cambridge, pp 278-287
- Lee DN (1976) A theory of visual control of braking based on information about time-to-collision. *Perception* 5: 437-459
- Lythgoe JN (1979) The ecology of vision. Clarendon Press, Oxford
- Marr D (1982) Vision: A computational investigation into the human representation and processing of visual information. Freeman and Co, San Francisco
- Metzger W (1975) Gesetze des Sehens. W Kramer Verlag, Frankfurt/Main
- Miles FA, Wallman J, eds (1993) Visual motion and its role in the stabilization of gaze. Elsevier, Amsterdam
- Mohn G, van Hof-van Duin J (1991) Development of spatial vision. In: Regan D (ed) Vision and visual dysfunction 10. Spatial vision. Macmillan Press, London, pp 179-211
- Nakayama K (1985) Biological image motion processing: A review. *Vision Res* 25: 625-660
- O'Carroll D, Bidwell NJ, Laughlin SB, Warrant EJ (1996) Insect motion detectors matched to visual ecology. *Nature* 382: 63-66, 1996
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381: 607-609
- Reichardt W (1987) Evaluation of optical motion information by movement detectors. *J Comp Physiol A* 161: 533-547
- Smith AT, Snowden RJ (1994) Visual detection of motion. Academic Press, London
- Srinivasan MV, Venkatesh S, eds (1997) From living eyes to seeing machines. Oxford University Press, Oxford
- Torre V, Poggio T (1978) A synaptic mechanism possibly underlying directional selectivity to motion. *Proc Roy Soc Lond B* 202: 409-416
- Ullman S (1979) The interpretation of visual motion. MIT Press, Cambridge
- Wade NJ (1998) A Natural history of vision. MIT Press, Cambridge MA
- Watanabe T (1998) High-level motion processing. Computational, neurobiological, and psychophysical perspectives. MIT Press, Cambridge