CHAPTER 1

Introduction

The topic of computer vision has produced many and varied texts over the last few years. One must ask at the outset, therefore, whether there is really any room for yet another sizeable book. In attempting to answer this question I have to admit that I am a relatively recent member of the computer vision community, having spent many years attempting to unravel the main mechanisms controlling human visual threshold performance. As a result of this latter background I have frequently felt unable to accept that the computer vision techniques widely discussed in the popular literature are remotely capable of achieving the levels of performance so readily achieved by human beings. It has been argued "Why should one *expect* computer vision to achieve similar performance to human beings?". On the other hand it has equally been argued "Why should computer vision be *limited* to the performance limits found to apply to human vision?". I have formed very firm convictions, from my wide studies of human vision and the visual environment, that there are very good reasons why *general* computer vision (as opposed to systems designed especially for certain very specific situations) *should* be limited roughly to the limits found in human vision. Perhaps the most compelling of such reasons is that a majority of tasks for which we attempt to employ computer vision are similar to visual tasks which have previously been carried out by human beings (and normally carried out very well!). At the same time I equally firmly believe that computer vision should be able to achieve performance close to those human limits with presently available optical and computer hardware. My main purpose in writing this book is, therefore, to present, under one cover, a wide variety of inter-related ideas and facts which have been collected together, both during my work on human visual performance limits and my more recent work in the field of computer vision. These collected ideas and facts together make what I consider to be a simple yet efficient and unified schema for the front end of a composite computer vision system. I make no apology for the fact that much of what I shall discuss *is* early processing, since I feel certain that, if one can extract the highest fidelity of a wide variety of fragmentary data by simple means, and if one can then process such fragmentary data simply, in order to provide transformed and still simple secondary data, then the task of 'perception' which is left should be relatively straight forward. I realise that this statement may seem rather presumptuous, but I shall hope to demonstrate by example what I mean by it as I proceed through the book. I shall also endeavour to address the 'perceptual' problem to some extent towards the end of the book.

I am very aware that, because my ideas quite often do not seem to be in line with popular or established teaching, many readers with prior knowledge in depth of some aspects of human vision and/or computer vision will find difficulty in equating with some of the ideas presented in early chapters. All I ask is that they read on, trying to keep an open mind, and only make judgement after studying the later chapters.

The sort of idea which has come out forcibly, in my studies on human vision, is that it is often the case that the processes used in human vision are far removed from the obvious which might be set up by engineers and scientists. For instance, optical and electro-optical engineers strive very hard with their imaging systems to provide as sharp an optical image as possible. How else, they argue, can one possibly expect to get good system performance? On the other hand the human eye is known to have relatively poor optics [see, for instance, 1.1, 1.2], yet the visual acuity is surprisingly good, whilst the so-called vernier acuity (that is, the ability to align two lines or edges as used in, for instance, vernier calipers), is truly spectacular (a small fraction of the resolution of the eye [1.3, 1.4]). Equally the human eyeball is known to exhibit very considerable tremor and drift, these being large compared to the spatial sampling interval or retinal receptor spacing [e.g. 1.5, 1.6]. These are just the sort of things which engineers try their hardest to stabilise out in their optical systems. Yet just such small amounts of motion, normally thought of as imperfections in engineering systems, are found to *improve* human visual performance rather than degrade it [e.g. 1.7, 1.8]. Yet again, such processes as stereo fusion and depth perception, remarkably effective even with certain forms of random dot pattern as used, for instance, by Julesz [1.9] and Rock [1.10], are carried out in human vision with effortless ease and, under ideal conditions, with a sensitivity approaching that of vernier acuity. Popular computer vision techniques to carry out the same type of process are usually both computationally slow, limited to single pixel accuracy and at best only a statistical approximation. I could go on – indeed as the book unfolds I intend to provide many more instances of where human vision seems to have utilised the unlikely or non-obvious to achieve quite remarkably simple, accurate and rapid solutions to visual problems.

Whilst I shall aim to *reference* appropriate popular treatments of various aspects of computer vision where appropriate, I must emphasise at the outset that I shall not *dwell* on any of these techniques. This is not from any wish to suppress them, but rather that there is a great deal to document on the specific treatment which I am trying to present. Therefore, to spend much time on other techniques would make the book prohibitively large. I feel sure that there are already adequate treatments of other techniques, indeed much more adequate than I could possibly aspire to producing. For *general* reading on more popular approaches I would recommend, in particular, [1.11] to [1.14].

My intention, then, is first to deal in some depth with a number of important facts and inferences about the human visual tract which are relevant to the approach to computer vision to be discussed (Chapter 2). Then I shall deal with what I consider to be the all important early extraction of basic fragmentary data (Chapters 3 and 4). I hope to show how this can be achieved very simply and elegantly by copying my interpretation of the basic structure of the human retina and early cortical processes following a unified approach. I shall then show how it is very easy to extract the fragmentary displacement data which are fundamental to the sensing of both motion and stereo disparity (Chapter 5), whilst in later chapters I shall provide copious illustration of how these fragmentary data can be readily processed to yield both local and global motion and stereo data (Chapters 9, 10 and 13). I shall subsequently discuss another very important aspect of human vision which has only recently been unravelled adequately (and is still in contention) – that of colour vision (Chapter 14). Here I shall show how, by very simple processes, it is possible to analyse *and* synthesise both luminance and chrominance distributions from conventional red rich, green rich and blue rich (R, G, B) input data. At the same time I shall illustrate how it is readily possible to provide a very important and special property of human colour vision – that is, the ability to sense *true* colour regardless of substantial variations in incident lighting.

All the foregoing are essentially early hard-wired processes in human vision. I shall supplement those discussions with a look at some of the higher level processes which are not necessarily currently positively located in human vision, but which can be predicted by compellingly simple and powerful operations on the basic fragmentary data already available from the early processes. Processes covered will include region segmentation based on energy *difference* data (Chapter 15), various forms of textural region segmentation (Chapter 16), local shape analysis (Chapter 12), a start to simple but positive shape synthesis etc.. Finally I shall address some *practical* aspects of computer vision systems and tasks (Chapter 17) and then attempt to speculate where we might be able to go from here, by following biological guidelines or by building on the wide range of simple, but high fidelity, data already demonstrated (Chapter 18).

An attempt has been made to cite major references from which ideas reported in the book have developed. However, in a number of cases the ideas have been triggered and developed over so many years, and from so many different sources, that it has proved difficult or impossible to trace the original source documents. In such cases I apologise to the authors of such papers for appearing not to acknowledge them, and to the present readership for not providing them with a full background of reading material. In addition, I am aware that, *since* the original ideas were developed, a wide variety of papers have been published which address many of the *individual* facets of vision included in my overall interpretation. By far the majority of those papers surveyed provide evidence broadly in support of my approach. Where appropriate I have tried to draw attention to such more recent references, although there appear to be so many available that I have certainly *not* surveyed them all. Again I apologise for all omissions in my reference citations.

It has been found, in our studies on *human* visual threshold performance, that a number of aspects of display imagery – particularly the image quality, display noise and discrete sampling – have marked effects on performance. Considerable space is therefore given over to assessment of such display characteristics, when applied to *computer* vision (Chapters 6, 7, and 8).

Much of the text and illustrations to be presented are based heavily on a large number of British Aerospace Reports prepared over the past 12 years. These reports have covered the growing development of our computer vision work, together with specialist studies into particular facets of the overall subject. In preparing the book, I have endeavoured to include sufficient details and illustrations such that the reader will not need, in general, to seek to refer to these source reports. Nevertheless, for completeness, I *have* included, at the end (see pp. 411–412), a short bibliography of various *open* publications which have appeared from time to time, covering both the computer vision studies and the earlier human visual performance and image evaluation studies. *Because* the computer vision studies have been in progress for several years, some aspects of the subject reported are already several years old, whilst other aspects are of very recent origin. This reflects what the author hopes is a self-coherent, progressive and unfolding story of a computer vision concept.

References

[1.1] Campbell F.W. and Gubisch R.W. (1966), 'Optical quality of the human eye', J. Physiol., *186*, 558.

[1.2] Overington I. (1976), 'Vision and Acquisition', Pentech Press, London. Chapter 2.3.

[1.3] LeGrand Y. (1967), 'Form and Space Vision', Indiana University Press, Bloomington and London.

[1.4] Berry R.N. (1948), 'Quantitative relation among vernier, real depth and stereoscopic depth acuities', J. Exp. Psychol., *38*, 708.

[1.5] Ditchburn R.W. and Foley-Fisher J.A. (1967), 'Assembled data on eye movements', Opt. Act., *14*, 113.

[1.6] Overington I. (1976), 'Vision and Acquisition', Pentech Press, London. Chapter 2.6.

[1.7] Overington I. (1976), 'Vision and Acquisition', Pentech Press, London. Chapter 4.10.

[1.8] Overington I. (1985), 'Sehfunktion und bilddarbietung', in 'Aspekte der Informationsverarbeitung' (Ed. H.R. Bodmann), Springer-Verlag, Heidelberg. (in German).

[1.9] Julesz B. (1971), 'Foundations of Cyclopean Perception', University of Chicago Press.

[1.10] Rock I. (1981), 'Anorthoscopic Perception', Scientific American, *244*, 103–110.

[1.11] Marr D. (1982), 'Vision', W.H. Freeman and Co., San Fransisco.

[1.12] Ballard D.H. and Brown C.M. (1982), 'Computer Vision', Prentice Hall.

[1.13] Braddick O.J. and Sleigh A.C. (Eds.) (1983), 'Physical and Biological Processing of Images', Springer-Verlag, Berlin, Heidelberg, New York.

[1.14] Gonzalez R.L. and Wintz P. (1987), 'Digital Image Processing', Addison-Wesley.