Preface

Attention in vision is something that I think has fascinated me since my undergraduate days at the University of Toronto. That is pretty surprising because I went to university wanting to be an aerospace engineer or maybe a physicist. In my first year, I subscribed to *Scientific American*, and in 1971 two papers caught my fancy: "Advances in Pattern Recognition" by R. Casey and G. Nagy and "Eye Movements and Visual Perception" by D. Noton and L. Stark. The first dealt in part with optical character recognition by computer, defining algorithms that might capture the process of vision and allow a computer to see. The second described the possible role of eye movements in vision and how they might define our internal representations of what we see. There had to be a connection! I have been trying to understand vision and what the connection between machine and biological vision might be since about 1974.

All through my graduate research, attention found its way into my work in some way. Back in the mid-1970s, there was a critical need for it in any large computer system: computing power was ridiculously meager by today's standards. I implemented my PhD thesis on a DEC PDP-11/45 with 256 kilobytes of memory! As a result, anything one could do to "focus" resources was a good thing. Similarly, if one looks at the computer vision research of the period (for that matter all of the artificial intelligence research, too), the inclusion of a "focus of attention" mechanism was not questioned.

But then in the early 1980s something happened, and, at least in the computational vision community, attention disappeared. I recall giving a seminar at a major U.S. university (nameless of course) where I spoke on my vision work, which included attention. I was taken aside by a very good friend afterward who apologized that many of the faculty did not attend my talk because, he said, they don't believe in attention at this school. I was surprised and disappointed, determined to "prove" that they were wrong. But how? These were really smart people, researchers for whom I had great respect. Could I really accomplish this? Maybe it was I who was mistaken? Within a couple of years of this event, 1985, as luck would have it, I

became part of an amazing organization, the Canadian Institute for Advanced Research. Its president, J. Fraser Mustard, believed that to tackle a difficult problem such as artificial intelligence, one really had to look at it from many perspectives: computation, engineering, neuroscience, psychology, philosophy, robotics, society, and more. It was this connection that appealed to me and that eventually led me to a path for approaching my goal. This superb collection of scientists from all these disciplines pushed me, and in my 10 years as a fellow of the institute, I learned more from them all than I can possibly acknowledge. The lessons were sometimes direct but most often indirect, absorbed simply by observation or through casual conversations. The most important lessons were abstracted from watching how the disciplines interacted with one another. Which was ready to absorb the results of the other? What were the barriers to communication? How does one transform theories from one domain into something useful for another? How could one convince one discipline that another had any utility for it? These, and more questions, made me think about how one might better conduct truly interdisciplinary research. Specifically, the perspectives of multiple disciplines became ingrained in me, and I eagerly embarked on trying to understand those different viewpoints and how they may complement and build on one another. The first papers from which the contents of this volume emerged were directly due to the influence of the Canadian Institute for Advanced Research and its Artificial Intelligence and Robotics program.

Looking at the field of computer vision or computational visual neuroscience today, attention is no longer invisible and seems to be playing an increasingly larger role. The push to develop models and systems that are biologically plausible is prominent. Still, attention is most often thought of as either selection of a region of interest to guide eye movements or as single-neuron modulation. Few seem interested in considering how these two processes might be related, and certainly not many seem interested in an overarching theory of attention.

Such a theory of attention, especially for vision, is what this book proposes, at least with respect to some of its foundations. Whether those foundations are successful in the long term will depend on how well their implications and predictions provide a basis for new insights into how the brain processes visual input and how well the resulting representations and computational constructs contribute to new computational vision systems. As with all scientific endeavors, time will tell.

The audience for which this book is intended is a broad and varied one, mirroring the diversity of research efforts into this domain. The book is intended not only for those embarking on research on visual attention and for its current practitioners but also for those who study vision more broadly, as it is central to the thesis of this volume that without attention, vision as we know it would not be possible. The list of interested disciplines is large: visual neuroscience, visual psychology, cognitive psychology, computational vision, computational neuroscience, engineering, com-

puter science, artificial intelligence, robotics, and more. It would be beyond the scope of any book to provide sufficient background so that anyone would find the book self-contained. To be sure, some background is presented in an abbreviated and certainly incomplete manner. Hopefully, enough pointers to relevant literature are included so the interested reader can track down what he or she might need. Those who have completed senior undergraduate or graduate-level courses in visual perception, computer vision, computational complexity, basic neuroanatomy of the visual system, and computational neuroscience will perhaps find the material more accessible than it will be to those who have not.

To provide a bit of assistance to some readers, the mathematical elements are confined to chapters 2 and 5 and appendixes B and C. Skipping these will of course lead to some gaps, but it shouldn't be too hard to follow the balance—unless you ask questions like "Why is he doing things this way?" In that case, you may have to simply bite the bullet and look at the math. Those who wish to see only the overview of the model can do so by reading chapters 4, 6, 7, and 8 and giving the early chapters less attention. For those who seek background on the main subject—visual attention—chapter 3 (and chapter 1 in a more general manner) is intended to be a comprehensive overview of attention theories and models. This literature is so large that gaps and unintentional omissions—for which I apologize—seem inevitable.

Those readers who identify with computer vision as their "home discipline" will undoubtedly be disappointed. But the current research directions in computer vision are not so compatible with the intent of this book. I am interested in using the language of computation, broadly speaking, to formalize and push forward our understanding of the mechanisms of vision and attention—both biological and artificial. Although I fully acknowledge the strong strides made by the computer vision community on the empirical and practical side of the discipline, that work is not covered in this book. Trust me, I may be more disappointed in this disconnect than you.

Many of the figures are better shown in color or as movies. There is a website associated with this book, <http://mitpress.mit.edu/Visual_Attention>, where one can see all color figures and movies. Where these are available, the citation in the book will be suffixed by "W." For example, if figure 7.3 has a color version, it can be found at the website as figure 7.3W, and it is referred to as such in this book. Movies are referred to as "movie 7.5W," not only pointing out that it is a movie but also that it is only available at the website. Although figures will be referred to with or without the "W" as appropriate, movies are only referred to with the "W" suffix.

Earlier, I wrote that two 1971 papers motivated my studies of vision and attention, but those were not my only motivation. My children played important roles, too, and it is for those roles that this book is dedicated to them. When my daughter, Lia (short for Ioulia), was born in 1985 (the same year that I joined the Canadian

Institute for Advanced Research, as I note in the preface—a fortuitous conjunction!), I was in the delivery room with my wife, Patty. I was the first to hold Lia on her birth and looked into her beautiful eyes—and was surprised! They did not seem to move in a coordinated manner; they gazed around apparently independently! The first thought in my head was, "What is going on in there to cause this? Is she okay?" After I was assured that there was nothing wrong, it occurred to me that I have to figure this out! Well I wound up not quite working on that side of the problem, but I do think this helped push me because the first paper that led to this book was written during the coming year. My son, Konstantine, was born in 1989, and this time I was better prepared for a birth, so no great surprises. However, about a year and a half later, he and I were lazing around at home on a Saturday morning looking for some cartoons on television to watch together. I found a program on robotics instead and was curious. It showed a disabled little boy operating a robotic toy-manipulation system. It was a very tedious system, and the juxtaposition of my healthy son playing on the floor beside me while watching the other little boy on television was actually painful to me. I thought that we should be able to do better, to build better systems to help. That was early 1991. My first paper on active vision was written as a result, appearing in 1992, and led to a robotic wheelchair project I named Playbot, intended to assist disabled children in play. So Lia and Konstantine, you were totally unaware of it at the time, but it is clear to me that if it weren't for you, my path would not be what it is today. And as I really like the research path that I am on, I thank you! You continue to inspire me every day with the wonder of how you are growing and becoming so much more than I will ever be.

My journey as a scientist has always had a modest goal. I have always viewed science as a race to solve a puzzle, a puzzle where the size, shape, and color of the pieces are unknown. Even the number of pieces and the eventual picture are unknown. Yet it is known that a picture exists, so we must discover what those puzzle pieces are and how they may fit together. My goal was always to be lucky enough to discover one or two of those puzzle pieces and to know where they fit within the full landscape that the puzzle represents. I think that every other scientist also has this as a goal. Only time will tell who discovers the right pieces for visual attention at the right time so that the picture is complete.