VISION

A Computational Investigation into the Human Representation and Processing of Visual Information

David Marr
Late of the Massachusetts Institute of Technology

W. H. Freeman and Company
San Francisco

Project Editor: Judith Wilson
Copy Editor: Paul Monsour
Production Coordinator: Linda Jupiter
Illustration Coordinator: Richard Quiñones
Designer: Ron Newcomer
Artists: Catherine Brandel and Victor Royer
Compositor: Graphic Typesetting Service
Printer and Binder: The Maple-Vail Book Manufacturing Group

Library of Congress Cataloging in Publication Data

Marr, David, 1945-1980.
Vision.

Bibliography: p.
Includes index.
1. Vision—Data processing. 2. Vision—Mathematical models. 3. Human information processing. I. Title.
QP475.M27 1982 152.1'4028'54 81-15076
ISBN 0-7167-1284-9

3 4 5 6 7 8 9 MP 1 0 8 9 8 7 6 5 4 3

To my parents and to Lucia

Note: Readers may require stereoscopic viewers in order to obtain the three-dimensional effects of the stereo images illustrated in this book. These viewers may be ordered from the following companies; please write to request current prices.

Hubbard Scientific Company
P.O. Box 104
Northbrook, Illinois 60062

Edmund Scientific Company
1776 Edscorp Building
Barrington, New Jersey 08007

The reader may be able to obtain the stereoscopic effect without an optical device: Hold the stereo image about ten inches away from the eyes and relax the eyes as if staring into the distance. Eventually the left-hand member of the pair seen by the right eye and the right-hand member of the pair seen by the left eye will merge to produce what will appear to be a three-dimensional image.

It will help to hold a fingertip about halfway between the stereo pair and your eyes. Adjust the position of the finger so that when looking with only your left eye, you see the finger in front of the right edge of the right-hand member of the pair. At the same time, when looking with your right eye only, try to see the finger in front of the right edge of the left-hand member of the pair. When your finger is so positioned, look at the finger with both eyes. This procedure will bring the two members of the stereo pair into registration, but they will be out of focus. Now relax your eyes and try to focus the stereo pair without losing the fixation on your finger. This trick seems to get easier as you get older.

Contents

Detailed Contents

PART II
VISION

Chapter 2
Representing the Image 41

Chapter 3
From Images to Surfaces 99

Preface

This book is meant to be enjoyed. It describes the adventures I have had in the years since Marvin Minsky and Seymour Papert invited me to the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology in 1973. Working conditions were ideal, thanks to Patrick Winston's skillful administration, to the generosity of the Advanced Research Projects Agency of the Department of Defense and of the National Science Foundation, and to the freedom arranged for me by Whitman Richards, under the benevolent eye of Richard Held. I was fortunate enough to meet and collaborate with a remarkable collection of people, most especially, Tomaso Poggio. Included among these people were many erstwhile students who became colleagues and from whom I learned much—Keith Nishihara, Shimon Ullman, Ken Forbus, Kent Stevens, Eric Grimson, Ellen Hildreth, Michael Riley, and John Batali. Berthold Horn kept us close to the physics of light, and Whitman Richards, to the abilities and inabilities of people.

In December 1977, certain events occurred that forced me to write this book a few years earlier than I had planned. Although the book has important gaps, which I hope will soon be filled, a new framework for studying vision is already clear and supported by enough solid results to be worth setting down as a coherent whole.

Many people have helped me to live through this somewhat difficult period. Particularly, my parents, my sister, my wife Lucia, and Jennifer, Tomaso, Shimon, Whitman, and Inge gave to me more than I often deserved; although mere thanks are inadequate, I thank them. William Prince steered me to Professor F. G. Hayhoe and Dr. John Rees at Addenbrooke's Hospital in Cambridge, and them I thank for giving me time.

Summer 1979

David Marr

should like to express our gratitude to those who helped us bring
id Marr's *Vision* to fulfillment.

thank Gunther Stent, whose friendship brought David Marr and W. H.
eman and Company together and whose sound guidance helped us
pare the book for publication.

thank David Marr's colleague, Keith Nishihara, for his skill and great
rt; the work could not have been finished without him.

thank David Marr's assistant, Carol Papineau, for attending so well to
needs of the manuscript and the publisher.

thank the vision group at the MIT Artificial Intelligence Laboratory,
ecially Ellen Hildreth and Eric Grimson, who participated in ways large
l small to bring this book to life.

The Publisher

PART I

Introduction and Philosophical Preliminaries

General Introduction

What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking. In other words, vision is the *process* of discovering from images what is present in the world, and where it is.

Vision is therefore, first and foremost, an information-processing task, but we cannot think of it just as a process. For if we are capable of knowing what is where in the world, our brains must somehow be capable of *representing* this information—in all its profusion of color and form, beauty, motion, and detail. The study of vision must therefore include not only the study of how to extract from images the various aspects of the world that are useful to us, but also an inquiry into the nature of the internal representations by which we capture this information and thus make it available as a basis for decisions about our thoughts and actions. This duality—the representation and the processing of information—lies at the heart of most information-processing tasks and will profoundly shape our investigation of the particular problems posed by vision.

The need to understand information-processing tasks and machines has arisen only quite recently. Until people began to dream of and then to build such machines, there was no very pressing need to think deeply

about them. Once people did begin to speculate about such tasks and machines, however, it soon became clear that many aspects of the world around us could benefit from an information-processing point of view. Most of the phenomena that are central to us as human beings—the mysteries of life and evolution, of perception and feeling and thought—are primarily phenomena of information processing, and if we are ever to understand them fully, our thinking about them must include this perspective.

The next point—which has to be made rather quickly to those who inhabit a world in which the local utility's billing computer is still capable of sending a final demand for $0.00—is to emphasize that saying that a job is "only" an information-processing task or that an organism is "only" an information-processing machine is not a limiting or a pejorative description. Even more importantly, I shall in no way use such a description to try to limit the kind of explanations that are necessary. Quite the contrary, in fact. One of the fascinating features of information-processing machines is that in order to understand them completely, one has to be satisfied with one's explanations at many different levels.

For example, let us look at the range of perspectives that must be satisfied before one can be said, from a human and scientific point of view, to have understood visual perception. First, and I think foremost, there is the perspective of the plain man. He knows what it is like to see, and unless the bones of one's arguments and theories roughly correspond to what this person knows to be true at first hand, one will probably be wrong (a point made with force and elegance by Austin, 1962). Second, there is the perspective of the brain scientists, the physiologists and anatomists who know a great deal about how the nervous system is built and how parts of it behave. The issues that concern them—how the cells are connected, why they respond as they do, the neuronal dogmas of Barlow (1972)—must be resolved and addressed in any full account of perception. And the same argument applies to the perspective of the experimental psychologists.

On the other hand, someone who has bought and played with a small home computer may make quite different demands. "If," he might say, "vision really is an information-processing task, then I should be able to make my computer do it, provided that it has sufficient power, memory, and some way of being connected to a home television camera." The explanation he wants is therefore a rather abstract one, telling him what to program and, if possible, a hint about the best algorithms for doing so. He doesn't want to know about rhodopsin, or the lateral geniculate nucleus, or inhibitory interneurons. He wants to know how to program vision.

The fundamental point is that in order to understand a device that performs an information-processing task, one needs many different kinds

of explanations. Part I of this book is concerned with this point, and it plays a prominent role because one of the keystones of the book is the realization that we have had to be more careful about what constitutes an explanation than has been necessary in other recent scientific developments, like those in molecular biology. For the subject of vision, there *is* no single equation or view that explains everything. Each problem has to be addressed from several points of view—as a problem in representing information, as a computation capable of deriving that representation, and as a problem in the architecture of a computer capable of carrying out both things quickly and reliably.

If one keeps strongly in mind this necessarily rather broad aspect of the nature of explanation, one can avoid a number of pitfalls. One consequence of an emphasis on information processing might be, for example, to introduce a comparison between the human brain and a computer. In a sense, of course, the brain is a computer, but to say this without qualification is misleading, because the essence of the brain is not simply that it is a computer but that it is a computer which is in the habit of performing some rather particular computations. The term *computer* usually refers to a machine with a rather standard type of instruction set that usually runs serially but nowadays sometimes in parallel, under the control of programs that have been stored in a memory. In order to understand such a computer, one needs to understand what it is made of, how it is put together, what its instruction set is, how much memory it has and how it is accessed, and how the machine may be made to run. But this forms only a small part of understanding a computer that is performing an information-processing task.

This point bears reflection, because it is central to why most analogies between brains and computers are too superficial to be useful. Think, for example, of the international network of airline reservation computers, which performs the task of assigning flights for millions of passengers all over the world. To understand this system it is not enough to know how a modern computer works. One also has to understand a little about what aircraft are and what they do; about geography, time zones, fares, exchange rates, and connections; and something about politics, diets, and the various other aspects of human nature that happen to be relevant to this particular task.

Thus the critical point is that understanding computers is different from understanding computations. To understand a computer, one has to study that computer. To understand an information-processing task, one has to study that information-processing task. To understand fully a particular machine carrying out a particular information-processing task, one has to do both things. Neither alone will suffice.

From a philosophical point of view, the approach that I describe is an extension of what have sometimes been called representational theories of mind. On the whole, it rejects the more recent excursions into the philosophy of perception, with their arguments about sense-data, the molecules of perception, and the validity of what the senses tell us; instead, this approach looks back to an older view, according to which the senses are for the most part concerned with telling one what is there. Modern representational theories conceive of the mind as having access to systems of internal representations; mental states are characterized by asserting what the internal representations currently specify, and mental processes by how such internal representations are obtained and how they interact.

This scheme affords a comfortable framework for our study of visual perception, and I am content to let it form the point of departure for our inquiry. As we shall see, pursuing this approach will lead us away from traditional avenues into what is almost a new intellectual landscape. Some of the things we find will seem strange, and it will be hard to reconcile subjectively some of the ideas and theories that are forced on us with what actually goes on inside ourselves when we open our eyes and look at things. Even the basic notion of what constitutes an explanation will have to be developed and broadened a little, to ensure that we do not leave anything out and that every important perspective on the problem is satisfied or satisfiable.

The book itself is divided into three parts. In the first are contained the philosophical preliminaries, a description of the approach, the representational framework that is proposed for the overall process of visual perception, and the way that led to it. I have adopted a fairly personal style in the hope that if the reader understands why particular directions were taken at each point, the reasons for the overall approach will be clearer.

The second part of the book, Chapters 2 to 6, contains the real analysis. It describes informally, but in some detail, how the approach and framework are actually realized, and the results that have been achieved.

The third part is somewhat unorthodox and consists of a set of questions and answers that are designed to help the reader to understand the way of thinking behind the approach—to help him acquire the right prejudices, if you like—and to relate these explanations to his personal experience of seeing. I have often found that one or two of the remarks set out in Part III have helped a person to see the point of part of the theory or to circumvent some private difficulty with it, and I hope they may serve a similar purpose here. The reader may find this section means more after having read the first two parts of the book, but an early glance at it may provide the motivation to take the trouble.

The detailed exposition comes, then, in Part II. Of course, the subject of human visual perception is not solved here by a long way. But over the last six years, my colleagues and I have been fortunate enough to see the establishment of an overall theoretical framework as well as the solution of several rather central problems in visual perception. We feel that the combination amounts to a reasonably strong case that the representational approach is a useful one, and the point of this book is to make that case. How far this approach can be pursued, of course, remains to be seen.