Contents

Preface

Machine vision is a young and rapidly changing field. It is exciting to write about it, but it is also hard to know when to stop, since new results appear at frequent intervals. This book grew out of notes for the undergraduate course 6.801, "Machine Vision," which I have taught at MIT for ten years. A draft version of the book has been in use for five years. The exercises are mostly from homework assignments and quizzes. The course is a "restricted elective," meaning that students can take it or choose instead some other course related to artificial intelligence. Most students who elect to take it do so in their junior year. Several chapters of the book have also been used in an intensive one-week summer course on robotics for people from industry and other universities.

Ten years ago, it was possible to introduce both robot manipulation and machine vision in one term. Knowledge in both areas advanced so quickly, however, that this is no longer feasible. (The other half of the original course has been expanded by Tomas Lozano-Pérez into 6.802, "Robot Manipulation.") In fact, a single term now seems too short even to talk about all of the interesting facets of machine vision.

Rapid progress in the field has made it possible to reduce the coverage of less significant areas. What is less significant is, to some extent, a matter of personal opinion, and this book reflects my preferences. It could not be otherwise, for an in-depth coverage of everything that has been done in machine vision would require much more space and would not lend itself to presentation under a coherent theme. Material that appears to lack solid theoretical foundations has been omitted.

Similarly, approaches that do not lead to useful methods for recovering information from images have been left out, even when they claim legitimacy by appeal to advanced mathematics. The book instead includes information that should be useful to engineers applying machine vision methods in the "real world." The chapters on binary image processing, for example, help explain and suggest how to improve the many commercial devices now available. The material on photometric stereo and the extended Gaussian image points the way to what may be the next thrust in commercialization of the results of research in this domain.

Implementation choices and specific algorithms are not always presented in full detail. Good implementations depend on particular features of available computing systems, and their presentation would tend to distract from the basic theme. Also, I believe that one should solve the basic machine vision problem before starting to worry about implementation. In most cases, implementation entails little more than straightforward application of the classic techniques of numerical analysis. Still, details relating to the efficient implementation in both software and hardware are included, for example, in the chapters on binary image processing.

Almost from the start, the course attracted graduate students who felt a need for exposure to the field, and some material has been included to exploit their greater mathematical sophistication. This material should be omitted in courses designed for a different audience, because it is hard to lay the mathematical foundations and simultaneously cover all the material in this book in a single term. This should present no difficulties, since several topics can be taught essentially independently from the rest. Also note that most of the necessary mathematical tools are developed in the book's appendix.

Aside from the obvious pairing of some of the chapters (3 and 4, 6 and 7, 10 and 11, 12 and 17, 16 and 18), there is in fact little interdependence among them. Students lacking background in linear systems theory may be better served if the chapters on image processing (6, 7, and perhaps 9) are omitted. Similarly, the chapters dealing with time-varying images (12 and 17) may also be left out without loss of continuity. A few chapters present material that is not as well developed as the rest, and if these are avoided also, one is left with a short, basic course consisting of chapters 1, 2, 3, 4, 10, 11, 16, and 18. There should be no problem covering that much in one term.

This book is intended to provide deep coverage of topics that I feel are reasonably well understood. This means that some topics are treated in less detail, and others, which I consider too ad hoc, not at all. In this regard the present book can be considered to be complementary to *Computer Vision* by Dana Ballard and Christopher Brown, a book which covers a larger number of topics, but in less depth. Also, many of the elementary concepts are dealt with in more detail in the second edition of *Digital Picture Processing* by Azriel Rosenfeld and Avinash Kak.

There is a strong connection between what is discussed here and the study of biological vision systems. I place less emphasis on this, using as an excuse the existence of an outstanding book on this topic, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, by the late David Marr. In a similar vein, I have given somewhat less prominence to work on edge detection, feature-based stereo, and some aspects of the interpretation of time-varying images, since the books *From Images to Surfaces: A Computational Study of the Human Early Visual System* by Eric Grimson, *The Measurement of Visual Motion* by Ellen Hildreth, and *The Interpretation of Visual Motion* by Shimon Ullman cover these subjects in greater detail than I can here. The same goes for pattern classification, given the existence of such classics as *Pattern Classification and Scene Analysis* by Richard Duda and Peter Hart. I have not been totally consistent, however, since there are two substantial chapters on image processing, despite the fact that the encyclopedic *Image Processing* by William Pratt covers this topic admirably. The reason is that this material is important for understanding preprocessing steps used in subsequent chapters.

Many of my students are at first surprised by the need for nontrivial mathematical techniques in this domain. To some extent this is because they have seen simple heuristic methods produce apparently startling results. Many such methods were discovered in the early days of machine vision and led to a false optimism about the expected rate of progress in the field. Later, significant limitations of these ad hoc approaches became apparent. It is obvious now that the study of machine vision needs to be supported by an understanding of image formation, just as the study of natural language requires some knowledge of linguistics. Not so long ago this would have been a minority view.

More seriously, machine vision is often considered merely as a means of analyzing sensor information for a system with "artificial intelligence." In artificial intelligence, there has been relatively little use, so far, of sophisticated mathematical manipulations. It is wrong to consider machine vision and robot manipulator control merely as the "I/O" of AI. The problems encountered in vision, manipulation, and locomotion are of interest in themselves. They are quite hard, and the tools required to attack them are nontrivial.

A system that successfully interacts with the environment can be understood, in part, by analyzing the physics of this interaction. In the case of vision this means that one ought to understand image formation if one wishes to recover information about the world from images. Modeling the physical interactions naturally leads to equations describing that interaction. The equations, in turn, suggest algorithms for recovering information about the three-dimensional world from images. This is my basic theme. Perhaps surprisingly, quite a few students find pleasure in applying mathematical methods learned in an abstract context to real problems. The material in this book provides them with motivation to practice methods and perhaps learn new concepts they would not have bothered with otherwise.

The emphasis in the first part of the book is on *early vision*—how to develop simple symbolic descriptions from images. Techniques for using these descriptions in spatial reasoning and in planning actions are less well developed and tend to depend on methodologies different from the ones appropriate to early vision. The last five chapters deal with methods that exploit the simple symbolic descriptions derived directly from images. Details of how to integrate a vision system into an overall robotics system are given in the final chapter, where a system for picking parts out of a bin is constructed.

A good part of the effort of writing this book went into the design of the exercises. They serve several purposes: Some help the reader practice ideas presented in a chapter; some develop ideas in more depth, perhaps using more sophisticated tools; and some introduce new research topics. The exercises in a given chapter are typically presented in this order, and hints are given to warn the reader about particularly difficult ones.

There has been a trend recently toward the use of more compact notation. In several instances, for example, components of vectors, such as surface normals and optical flow velocities, were used in early work. The current tendency is to use the vectors directly; for example, the Gaussian sphere is now employed instead of gradient space in the specification of surface orientation. In the body of this text I have used the component notation, which is easier to grasp initially. In some of the exercises, however, I have tried to show how problems can be solved more efficiently using more compact notation.

Most of the material included in this book has been presented elsewhere, but here it is organized in a more coherent way, using a consistent notation. In a few instances, new methods are presented that have not been published before. Because the field is changing rapidly, some of what is presented here may become obsolete, or at least of less interest, in just a few years. Conversely, some things that I have not covered may eventually form the basis for exciting new results. This is not a serious shortcoming, however, since my concern is more with the development of a solid approach to research in machine vision than with specific techniques for tackling particular problems.

When will we have a "general-purpose" vision system? Not in the foreseeable future, is my answer. This is not to say that machine vision is merely an intellectual exercise of no practical import. On the contrary, tremendous progress has been made in two ways: (a) by concentrating on a particular aspect of vision, such as interpretation of stereo pairs, and (b) by concentrating on a particular application, such as the alignment of parts for automated assembly. A truly general-purpose vision system would have to deal with all aspects of vision and be applicable to all problems that can be solved using visual information. Among other things, it would have to be able to reason about the physical world.

B.K.P.H.

Acknowledgments

The students of the "Machine Vision" course at MIT deserve much credit for helping me to formulate and revise this material. My teaching assistants have contributed to the generation of many of the problems. Robert Sjoberg also prepared careful notes on several topics that I have unfortunately been unable to incorporate due to time pressure. Robert Sjoberg, Andy Moulton, Eric Bier, Michael Gennert, and Jazek Myczkowski provided numerous useful comments on earlier drafts.

A few of the chapters are based on papers that I wrote jointly with others. I would like to thank Michael Brooks for his contribution to the discussion of the shape-from-shading problem (chapter 11), Brian Schunck for his help in the development of methods for the analysis of optical flow (chapter 12), Anna Bruss for her contribution to the analysis of the passive navigation problem (chapter 17), and Katsushi Ikeuchi for his serious dedication to implementing the bin-picking system (chapter 18).

Christopher Brown, Herbert Freeman, Eric Grimson, Ramesh Jain, Alan Mackworth, and Lothar Rossol provided helpful comments on an early draft. Michael Brady, Michael Brooks, Michael Gennert, and Ellen Hildreth went over recent versions of the book and made many useful suggestions. Michael Gennert contributed the problems on pattern classification (chapter 14). Careful reading by Boris Katz and Larry Cohen helped eliminate the more blatant linguistic problems. Unfortunately, I could not resist the temptation to rewrite much of the material as time went on and so have, no doubt, in the process reintroduced many bugs and typos.

The Department of Electrical Engineering and Computer Science gave me a six-month sabbatical to write the first draft. Carol Roberts typed that draft. Blythe Heepe drew most of the figures. Phyllis Rogers helped me with the bibliography. Michael Gennert came to my rescue by taking over final preparation of the camera ready copy. The illustration on the dust cover is used with the kind permission of the artist, Hajime Sorayama.

Marvin Minsky got me started in machine vision by suggesting the recovery of shape from brightness gradations in an image as a thesis topic. Patrick Winston was supportive of my approach to machine vision from the start, when it was not a popular one. Marvin is responsible for the creation, and Patrick for the survival and expansion of the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology, where work on machine vision has flourished for almost twenty years.