COMPUTER VISION
A MODERN APPROACH

SECOND EDITION

DAVID A. FORSYTH
University of Illinois at Urbana-Champaign

JEAN PONCE
Ecole Normale Supérieure

PEARSON

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Vice President and Editorial Director, ECS:
Marcia Horton
Editor in Chief: Michael Hirsch
Executive Editor: Tracy Dunkelberger
Senior Project Manager: Carole Snyder
Vice President Marketing: Patrice Jones
Marketing Manager: Yez Alayan
Marketing Coordinator: Kathryn Ferranti
Marketing Assistant: Emma Snider
Vice President and Director of Production:
Vince O'Brien
Managing Editor: Jeff Holcomb
Senior Production Project Manager: Marilyn Lloyd
Senior Operations Supervisor: Alan Fischer
Operations Specialist: Lisa McDowell
Art Director, Cover: Jayne Conte
Text Permissions: Dana Weightman/RightsHouse,
Inc. and Jen Roach/PreMediaGlobal
Cover Image: © Maxppp/ZUMAPRESS.com
Media Editor: Dan Sandin
Composition: David Forsyth
Printer/Binder: Edwards Brothers
Cover Printer: Lehigh-Phoenix Color

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on the appropriate page within text.

Library of Congress Cataloging-in-Publication Data available upon request

10 9 8 7 6 5 4 3 2 1

PEARSON

ISBN-13: 978-0-13-608592-8
ISBN-10: 0-13-608592-X

To my family—DAF

To my father, Jean-Jacques Ponce —JP

Contents

Preface

Computer vision as a field is an intellectual frontier. Like any frontier, it is exciting and disorganized, and there is often no reliable authority to appeal to. Many useful ideas have no theoretical grounding, and some theories are useless in practice; developed areas are widely scattered, and often one looks completely inaccessible from the other. Nevertheless, we have attempted in this book to present a fairly orderly picture of the field.

We see computer vision—or just "vision"; apologies to those who study human or animal vision—as an enterprise that uses statistical methods to disentangle data using models constructed with the aid of geometry, physics, and learning theory. Thus, in our view, vision relies on a solid understanding of cameras and of the physical process of image formation (Part I of this book) to obtain simple inferences from individual pixel values (Part II), combine the information available in multiple images into a coherent whole (Part III), impose some order on groups of pixels to separate them from each other or infer shape information (Part IV), and recognize objects using geometric information or probabilistic techniques (Part V). Computer vision has a wide variety of applications, both old (e.g., mobile robot navigation, industrial inspection, and military intelligence) and new (e.g., human computer interaction, image retrieval in digital libraries, medical image analysis, and the realistic rendering of synthetic scenes in computer graphics). We discuss some of these applications in part VII.

IN THE SECOND EDITION

We have made a variety of changes since the first edition, which we hope have improved the usefulness of this book. Perhaps the most important change follows a big change in the discipline since the last edition. Code and data are now widely published over the Internet. It is now quite usual to build systems out of other people's published code, at least in the first instance, and to evaluate them on other people's datasets. In the chapters, we have provided guides to experimental resources available online. As is the nature of the Internet, not all of these URL's will work all the time; we have tried to give enough information so that searching Google with the authors' names or the name of the dataset or codes will get the right result.

Other changes include:

- We have **simplified.** We give a simpler, clearer treatment of mathematical topics. We have particularly simplified our treatment of cameras (Chapter 1), shading (Chapter 2), and reconstruction from two views (Chapter 7) and from multiple views (Chapter 8)
- We describe a **broad range of applications**, including image-based modelling and rendering (Chapter 19), image search (Chapter 22), building image mosaics (Section 12.1), medical image registration (Section 12.3), interpreting range data (Chapter 14), and understanding human activity (Chapter 21).

- We have written a comprehensive treatment of the **modern features**, particularly HOG and SIFT (both in Chapter 5), that drive applications ranging from building image mosaics to object recognition.
- We give a detailed treatment of **modern image editing techniques**, including removing shadows (Section 3.5), filling holes in images (Section 6.3), noise removal (Section 6.4), and interactive image segmentation (Section 9.2).
- We give a comprehensive treatment of **modern object recognition techniques**. We start with a practical discussion of classifiers (Chapter 15); we then describe standard methods for image classification techniques (Chapter 16), and object detection (Chapter 17). Finally, Chapter 18 reviews a wide range of recent topics in object recognition.
- Finally, this book has a very detailed index, and a bibliography that is as comprehensive and up-to-date as we could make it.

WHY STUDY VISION?

Computer vision's great trick is extracting descriptions of the world from pictures or sequences of pictures. This is unequivocally useful. Taking pictures is usually nondestructive and sometimes discreet. It is also easy and (now) cheap. The descriptions that users seek can differ widely between applications. For example, a technique known as structure from motion makes it possible to extract a representation of what is depicted and how the camera moved from a series of pictures. People in the entertainment industry use these techniques to build three-dimensional (3D) computer models of buildings, typically keeping the structure and throwing away the motion. These models are used where real buildings cannot be; they are set fire to, blown up, etc. Good, simple, accurate, and convincing models can be built from quite small sets of photographs. People who wish to control mobile robots usually keep the motion and throw away the structure. This is because they generally know something about the area where the robot is working, but usually don't know the precise robot location in that area. They can determine it from information about how a camera bolted to the robot is moving.

There are a number of other, important applications of computer vision. One is in medical imaging: one builds software systems that can enhance imagery, or identify important phenomena or events, or visualize information obtained by imaging. Another is in inspection: one takes pictures of objects to determine whether they are within specification. A third is in interpreting satellite images, both for military purposes (a program might be required to determine what militarily interesting phenomena have occurred in a given region recently; or what damage was caused by a bombing) and for civilian purposes (what will this year's maize crop be? How much rainforest is left?) A fourth is in organizing and structuring collections of pictures. We know how to search and browse text libraries (though this is a subject that still has difficult open questions) but don't really know what to do with image or video libraries.

Computer vision is at an extraordinary point in its development. The subject itself has been around since the 1960s, but only recently has it been possible to build useful computer systems using ideas from computer vision. This flourishing has been driven by several trends: Computers and imaging systems have become very cheap. Not all that long ago, it took tens of thousands of dollars to get good digital color images; now it takes a few hundred at most. Not all that long ago, a color printer was something one found in few, if any, research labs; now they are in many homes. This means it is easier to do research. It also means that there are many people with problems to which the methods of computer vision apply. For example, people would like to organize their collections of photographs, make 3D models of the world around them, and manage and edit collections of videos. Our understanding of the basic geometry and physics underlying vision and, more important, what to do about it, has improved significantly. We are beginning to be able to solve problems that lots of people care about, but none of the hard problems have been solved, and there are plenty of easy ones that have not been solved either (to keep one intellectually fit while trying to solve hard problems). It is a great time to be studying this subject.

What Is in this Book

This book covers what we feel a computer vision professional ought to know. However, it is addressed to a wider audience. We hope that those engaged in computational geometry, computer graphics, image processing, imaging in general, and robotics will find it an informative reference. We have tried to make the book accessible to senior undergraduates or graduate students with a passing interest in vision. Each chapter covers a different part of the subject, and, as a glance at Table 1 will confirm, chapters are relatively independent. This means that one can dip into the book as well as read it from cover to cover. Generally, we have tried to make chapters run from easy material at the start to more arcane matters at the end. Each chapter has brief notes at the end, containing historical material and assorted opinions. We have tried to produce a book that describes ideas that are useful, or likely to be so in the future. We have put emphasis on understanding the basic geometry and physics of imaging, but have tried to link this with actual applications. In general, this book reflects the enormous recent influence of geometry and various forms of applied statistics on computer vision.

Reading this Book

A reader who goes from cover to cover will hopefully be well informed, if exhausted; there is too much in this book to cover in a one-semester class. Of course, prospective (or active) computer vision professionals should read every word, do all the exercises, and report any bugs found for the third edition (of which it is probably a good idea to plan on buying a copy!). Although the study of computer vision does not require deep mathematics, it does require facility with a lot of different mathematical ideas. We have tried to make the book self-contained, in the sense that readers with the level of mathematical sophistication of an engineering senior should be comfortable with the material of the book and should not need to refer to other texts. We have also tried to keep the mathematics to the necessary minimum—after all, this book is about computer vision, not applied mathematics—and have chosen to insert what mathematics we have kept in the main chapter bodies instead of a separate appendix.

TABLE 1: Dependencies between chapters: It will be difficult to read a chapter if you don't have a good grasp of the material in the chapters it "requires." If you have not read the chapters labeled "helpful," you might need to look up one or two things.

Part		Chapter	Requires	Helpful
I	1:	Geometric Camera Models		
	2:	Light and Shading		
	3:	Color	2	
II	4:	Linear Filters		
	5:	Local Image Features	4	
	6:	Texture	5, 4	2
III	7:	Stereopsis	1	22
	8:	Structure from Motion	1, 7	22
IV	9:	Segmentation by Clustering		2, 3, 4, 5, 6, 22
	10:	Grouping and Model Fitting		9
	11:	Tracking		2, 5, 22
V	12:	Registration	1	14
	13:	Smooth Surfaces and Their Outlines	1	
	14:	Range Data		12
	15:	Learning to Classify		22
	16:	Classifying Images	15, 5	
	17:	Detecting Objects in Images	16, 15, 5	
	18:	Topics in Object Recognition	17, 16, 15, 5	
VI	19:	Image-Based Modeling and Rendering	1, 2, 7, 8	
	20:	Looking at People		17, 16, 15, 11, 5
	21:	Image Search and Retrieval		17, 16, 15, 11, 5
VII	22:	Optimization Techniques		

Generally, we have tried to reduce the interdependence between chapters, so that readers interested in particular topics can avoid wading through the whole book. It is not possible to make each chapter entirely self-contained, however, and Table 1 indicates the dependencies between chapters.

We have tried to make the index comprehensive, so that if you encounter a new term, you are likely to find it in the book by looking it up in the index. Computer vision is now fortunate in having a rich range of intellectual resources. Software and datasets are widely shared, and we have given pointers to useful datasets and software in relevant chapters; you can also look in the index, under "software" and under "datasets," or under the general topic.

We have tried to make the bibliography comprehensive, without being overwhelming. However, we have not been able to give complete bibliographic references for any topic, because the literature is so large.

What Is Not in this Book

The computer vision literature is vast, and it was not easy to produce a book about computer vision that could be lifted by ordinary mortals. To do so, we had to cut material, ignore topics, and so on.

We left out some topics because of personal taste, or because we became exhausted and stopped writing about a particular area, or because we learned about them too late to put them in, or because we had to shorten some chapter, or because we didn't understand them, or any of hundreds of other reasons. We have tended to omit detailed discussions of material that is mainly of historical interest, and offer instead some historical remarks at the end of each chapter.

We have tried to be both generous and careful in attributing ideas, but neither of us claims to be a fluent intellectual archaeologist, and computer vision is a very big topic indeed. This means that some ideas may have deeper histories than we have indicated, and that we may have omitted citations.

There are several recent textbooks on computer vision. Szeliski (2010) deals with the whole of vision. Parker (2010) deals specifically with algorithms. Davies (2005) and Steger *et al.* (2008) deal with practical applications, particularly registration. Bradski and Kaehler (2008) is an introduction to OpenCV, an important open-source package of computer vision routines.

There are numerous more specialized references. Hartley and Zisserman (2000*a*) is a comprehensive account of what is known about multiple view geometry and estimation of multiple view parameters. Ma *et al.* (2003*b*) deals with 3D reconstruction methods. Cyganek and Siebert (2009) covers 3D reconstruction and matching. Paragios *et al.* (2010) deals with mathematical models in computer vision. Blake *et al.* (2011) is a recent summary of what is known about Markov random field models in computer vision. Li and Jain (2005) is a comprehensive account of face recognition. Moeslund *et al.* (2011), which is in press at time of writing, promises to be a comprehensive account of computer vision methods for watching people. Dickinson *et al.* (2009) is a collection of recent summaries of the state of the art in object recognition. Radke (2012) is a forthcoming account of computer vision methods applied to special effects.

Much of computer vision literature appears in the proceedings of various conferences. The three main conferences are: the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); the IEEE International Conference on Computer Vision (ICCV); and the European Conference on Computer Vision. A significant fraction of the literature appears in regional conferences, particularly the Asian Conference on Computer Vision (ACCV) and the British Machine Vision Conference (BMVC). A high percentage of published papers are available on the web, and can be found with search engines; while some papers are confined to pay-libraries, to which many universities provide access, most can be found without cost.

ACKNOWLEDGMENTS

In preparing this book, we have accumulated a significant set of debts. A number of anonymous reviewers read several drafts of the book for both first and second edition and made extremely helpful contributions. We are grateful to them for their time and efforts.

Our editor for the first edition, Alan Apt, organized these reviews with the

help of Jake Warde. We thank them both. Leslie Galen, Joe Albrecht, and Dianne Parish, of Integre Technical Publishing, helped us overcome numerous issues with proofreading and illustrations in the first edition.

Our editor for the second edition, Tracy Dunkelberger, organized reviews with the help of Carole Snyder. We thank them both. We thank Marilyn Lloyd for helping us get over various production problems.

Both the overall coverage of topics and several chapters were reviewed by various colleagues, who made valuable and detailed suggestions for their revision. We thank Narendra Ahuja, Francis Bach, Kobus Barnard, Margaret Fleck, Martial Hebert, Julia Hockenmaier, Derek Hoiem, David Kriegman, Jitendra Malik, and Andrew Zisserman.

A number of people contributed suggestions, ideas for figures, proofreading comments, and other valuable material, while they were our students. We thank Okan Arikan, Louise Benoît, Tamara Berg, Sébastien Blind, Y-Lan Boureau, Liang-Liang Cao, Martha Cepeda, Stephen Chenney, Frank Cho, Florent Couzinie-Devy, Olivier Duchenne, Pinar Duygulu, Ian Endres, Ali Farhadi, Yasutaka Furukawa, Yakup Genc, John Haddon, Varsha Hedau, Nazli Ikizler-Cinbis, Leslie Ikemoto, Sergey Ioffe, Armand Joulin, Kevin Karsch, Svetlana Lazebnik, Cathy Lee, Binbin Liao, Nicolas Loeff, Julien Mairal, Sung-il Pae, David Parks, Deva Ramanan, Fred Rothganger, Amin Sadeghi, Alex Sorokin, Attawith Sudsang, Du Tran, Duan Tran, Gang Wang, Yang Wang, Ryan White, and the students in several offerings of our vision classes at UIUC, U.C. Berkeley and ENS.

We have been very lucky to have colleagues at various universities use (often rough) drafts of our book in their vision classes. Institutions whose students suffered through these drafts include, in addition to ours, Carnegie-Mellon University, Stanford University, the University of Wisconsin at Madison, the University of California at Santa Barbara and the University of Southern California; there may be others we are not aware of. We are grateful for all the helpful comments from adopters, in particular Chris Bregler, Chuck Dyer, Martial Hebert, David Kriegman, B.S. Manjunath, and Ram Nevatia, who sent us many detailed and helpful comments and corrections.

The book has also benefitted from comments and corrections from Karteek Alahari, Aydin Alaylioglu, Srinivas Akella, Francis Bach, Marie Banich, Serge Belongie, Tamara Berg, Ajit M. Chaudhari, Navneet Dalal, Jennifer Evans, Yasutaka Furukawa, Richard Hartley, Glenn Healey, Mike Heath, Martial Hebert, Janne Heikkilä, Hayley Iben, Stéphanie Jonquières, Ivan Laptev, Christine Laubenberger, Svetlana Lazebnik, Yann LeCun, Tony Lewis, Benson Limketkai, Julien Mairal, Simon Maskell, Brian Milch, Roger Mohr, Deva Ramanan, Guillermo Sapiro, Cordelia Schmid, Brigitte Serlin, Gerry Serlin, Ilan Shimshoni, Jamie Shotton, Josef Sivic, Eric de Sturler, Camillo J. Taylor, Jeff Thompson, Claire Vallat, Daniel S. Wilkerson, Jinghan Yu, Hao Zhang, Zhengyou Zhang, and Andrew Zisserman.

In the first edition, we said

> If you find an apparent typographic error, please email DAF... with the details, using the phrase "book typo" in your email; we will try to credit the first finder of each typo in the second edition.

which turns out to have been a mistake. DAF's ability to manage and preserve email logs was just not up to this challenge. We thank all finders of typographic errors; we have tried to fix the errors and have made efforts to credit all the people who have helped us.

We also thank P. Besl, B. Boufama, J. Costeira, P. Debevec, O. Faugeras, Y. Genc, M. Hebert, D. Huber, K. Ikeuchi, A.E. Johnson, T. Kanade, K. Kutulakos, M. Levoy, Y. LeCun, S. Mahamud, R. Mohr, H. Moravec, H. Murase, Y. Ohta, M. Okutami, M. Pollefeys, H. Saito, C. Schmid, J. Shotton, S. Sullivan, C. Tomasi, and M. Turk for providing the originals of some of the figures shown in this book.

DAF acknowledges ongoing research support from the National Science Foundation. Awards that have directly contributed to the writing of this book are IIS-0803603, IIS-1029035, and IIS-0916014; other awards have shaped the view described here. DAF acknowledges ongoing research support from the Office of Naval Research, under awards N00014-01-1-0890 and N00014-10-1-0934, which are part of the MURI program. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of NSF or ONR.

DAF acknowledges a wide range of intellectual debts, starting at kindergarten. Important figures in the very long list of his creditors include Gerald Alanthwaite, Mike Brady, Tom Fair, Margaret Fleck, Jitendra Malik, Joe Mundy, Mike Rodd, Charlie Rothwell, and Andrew Zisserman. JP cannot even remember kindergarten, but acknowledges his debts to Olivier Faugeras, Mike Brady, and Tom Binford. He also wishes to thank Sharon Collins for her help. Without her, this book, like most of his work, probably would have never been finished. Both authors would also like to acknowledge the profound influence of Jan Koenderink's writings on their work at large and on this book in particular.

Figures: Some images used herein were obtained from IMSI's Master Photos Collection, 1895 Francisco Blvd. East, San Rafael, CA 94901-5506, USA. We have made extensive use of figures from the published literature; these figures are credited in their captions. We thank the copyright holders for extending permission to use these figures.

Bibliography: In preparing the bibliography, we have made extensive use of Keith Price's excellent computer vision bibliography, which can be found at `http://iris.usc.edu/Vision-Notes/bibliography/contents.html`.

TABLE 2: A one-semester introductory class in computer vision for seniors or first-year graduate students in computer science, electrical engineering, or other engineering or science disciplines.

Week	Chapter	Sections	Key topics
1	1, 2	1.1, 2.1, 2.2.x	pinhole cameras, pixel shading models, one inference from shading example
2	3	3.1–3.5	human color perception, color physics, color spaces, image color model
3	4	all	linear filters
4	5	all	building local features
5	6	6.1, 6.2	texture representations from filters, from vector quantization
6	7	7.1, 7.2	binocular geometry, stereopsis
7	8	8.1	structure from motion with perspective cameras
8	9	9.1–9.3	segmentation ideas, applications, segmentation by clustering pixels
9	10	10.1–10.4	Hough transform, fitting lines, robustness, RANSAC,
10	11	11.1-11.3	simple tracking strategies, tracking by matching, Kalman filters, data association
11	12	all	registration
12	15	all	classification
13	16	all	classifying images
14	17	all	detection
15	choice	all	one of chapters 14, 19, 20, 21 (application topics)

SAMPLE SYLLABUSES

The whole book can be covered in two (rather intense) semesters, by starting at the first page and plunging on. Ideally, one would cover one application chapter—probably the chapter on image-based rendering—in the first semester, and the other one in the second. Few departments will experience heavy demand for such a detailed sequence of courses. We have tried to structure this book so that instructors can choose areas according to taste. Sample syllabuses for busy 15-week semesters appear in Tables 2 to 6, structured according to needs that can reasonably be expected. We would encourage (and expect!) instructors to rearrange these according to taste.

Table 2 contains a suggested syllabus for a one-semester introductory class in computer vision for seniors or first-year graduate students in computer science, electrical engineering, or other engineering or science disciplines. The students receive a broad presentation of the field, including application areas such as digital libraries and image-based rendering. Although the hardest theoretical material is omitted, there is a thorough treatment of the basic geometry and physics of image formation. We assume that students will have a wide range of backgrounds, and can be assigned background readings in probability. We have put off the application chapters to the end, but many may prefer to cover them earlier.

Table 3 contains a syllabus for students of computer graphics who want to know the elements of vision that are relevant to their topic. We have emphasized methods that make it possible to recover object models from image information;

TABLE 3: A syllabus for students of computer graphics who want to know the elements of vision that are relevant to their topic.

Week	Chapter	Sections	Key topics
1	1, 2	1.1, 2.1, 2.2.4	pinhole cameras, pixel shading models, photometric stereo
2	3	3.1–3.5	human color perception, color physics, color spaces, image color model
3	4	all	linear filters
4	5	all	building local features
5	6	6.3, 6.4	texture synthesis, image denoising
6	7	7.1, 7.2	binocular geometry, stereopsis
7	7	7.4, 7.5	advanced stereo methods
8	8	8.1	structure from motion with perspective cameras
9	10	10.1–10.4	Hough transform, fitting lines, robustness, RANSAC,
10	9	9.1–9.3	segmentation ideas, applications, segmentation by clustering pixels
11	11	11.1-11.3	simple tracking strategies, tracking by matching, Kalman filters, data association
12	12	all	registration
13	14	all	range data
14	19	all	image-based modeling and rendering
15	13	all	surfaces and outlines

understanding these topics needs a working knowledge of cameras and filters. Tracking is becoming useful in the graphics world, where it is particularly important for motion capture. We assume that students will have a wide range of backgrounds, and have some exposure to probability.

Table 4 shows a syllabus for students who are primarily interested in the applications of computer vision. We cover material of most immediate practical interest. We assume that students will have a wide range of backgrounds, and can be assigned background reading.

Table 5 is a suggested syllabus for students of cognitive science or artificial intelligence who want a basic outline of the important notions of computer vision. This syllabus is less aggressively paced, and assumes less mathematical experience.

Our experience of teaching computer vision is that no single idea presents any particular conceptual difficulties, though some are harder than others. Difficulties are caused by the tremendous number of new ideas required by the subject. Each subproblem seems to require its own way of thinking, and new tools to cope with it. This makes learning the subject rather daunting. Table 6 shows a sample syllabus for students who are really not bothered by these difficulties. They would need to have quite a strong interest in applied mathematics, electrical engineering or physics, and be very good at picking things up as they go along. This syllabus sets a furious pace, and assumes that students can cope with a lot of new material.

NOTATION

We use the following notation throughout the book: Points, lines, and planes are denoted by Roman or Greek letters in italic font (e.g., P, Δ, or Π). Vectors are

TABLE 4: A syllabus for students who are primarily interested in the applications of computer vision.

Week	Chapter	Sections	Key topics
1	1, 2	1.1, 2.1, 2.2.4	pinhole cameras, pixel shading models, photometric stereo
2	3	3.1–3.5	human color perception, color physics, color spaces, image color model
3	4	all	linear filters
4	5	all	building local features
5	6	6.3, 6.4	texture synthesis, image denoising
6	7	7.1, 7.2	binocular geometry, stereopsis
7	7	7.4, 7.5	advanced stereo methods
8	8, 9	8.1, 9.1–9.2	structure from motion with perspective cameras, segmentation ideas, applications
9	10	10.1–10.4	Hough transform, fitting lines, robustness, RANSAC,
10	12	all	registration
11	14	all	range data
12	16	all	classifying images
13	19	all	image based modeling and rendering
14	20	all	looking at people
15	21	all	image search and retrieval

usually denoted by Roman or Greek bold-italic letters (e.g., $\boldsymbol{v}$, $\boldsymbol{P}$, or $\boldsymbol{\xi}$), but the vector joining two points P and Q is often denoted by $\overrightarrow{PQ}$. Lower-case letters are normally used to denote geometric figures in the image plane (e.g., p, $\boldsymbol{p}$, δ), and upper-case letters are used for scene objects (e.g., P, Π). Matrices are denoted by Roman letters in calligraphic font (e.g., $\mathcal{U}$).

The familiar three-dimensional Euclidean space is denoted by $\mathbb{E}^3$, and the vector space formed by n-tuples of real numbers with the usual laws of addition and multiplication by a scalar is denoted by $\mathbb{R}^n$, with $\mathbf{0}$ being used to denote the zero vector. Likewise, the vector space formed by $m \times n$ matrices with real entries is denoted by $\mathbb{R}^{m\times n}$. When $m = n$, Id is used to denote the identity matrix—that is, the $n \times n$ matrix whose diagonal entries are equal to 1 and nondiagonal entries are equal to 0. The transpose of the $m \times n$ matrix $\mathcal{U}$ with coefficients u_{ij} is the $n \times m$ matrix denoted by $\mathcal{U}^T$ with coefficients u_{ji}. Elements of $\mathbb{R}^n$ are often identified with column vectors or $n \times 1$ matrices, for example, $\boldsymbol{a} = (a_1, a_2, a_3)^T$ is the transpose of a 1×3 matrix (or *row vector*), i.e., an 3×1 matrix (or *column vector*), or equivalently an element of $\mathbb{R}^3$.

The *dot product* (or *inner product*) of two vectors $\boldsymbol{a} = (a_1, \dots, a_n)^T$ and $\boldsymbol{b} = (b_1, \dots, b_n)^T$ in $\mathbb{R}^n$ is defined by

$$\boldsymbol{a} \cdot \boldsymbol{b} = a_1 b_1 + \cdots + a_n b_n,$$

and it can also be written as a matrix product, i.e., $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{a}^T \boldsymbol{b} = \boldsymbol{b}^T \boldsymbol{a}$. We denote by $|\boldsymbol{a}|^2 = \boldsymbol{a} \cdot \boldsymbol{a}$ the square of the Euclidean norm of the vector $\boldsymbol{a}$ and denote by d the distance function induced by the Euclidean norm in $\mathbb{E}^n$, i.e., $d(P, Q) = |\overrightarrow{PQ}|$. Given a matrix $\mathcal{U}$ in $\mathbb{R}^{m\times n}$, we generally use $|U|$ to denote its *Frobenius norm*, i.e., the square root of the sum of its squared entries.

TABLE 5: For students of cognitive science or artificial intelligence who want a basic outline of the important notions of computer vision.

Week	Chapter	Sections	Key topics
1	1, 2	1.1, 2.1, 2.2.x	pinhole cameras, pixel shading models, one inference from shading example
2	3	3.1–3.5	human color perception, color physics, color spaces, image color model
3	4	all	linear filters
4	5	all	building local features
5	6	6.1, 6.2	texture representations from filters, from vector quantization
6	7	7.1, 7.2	binocular geometry, stereopsis
8	9	9.1–9.3	segmentation ideas, applications, segmentation by clustering pixels
9	11	11.1, 11.2	simple tracking strategies, tracking using matching, optical flow
10	15	all	classification
11	16	all	classifying images
12	20	all	looking at people
13	21	all	image search and retrieval
14	17	all	detection
15	18	all	topics in object recognition

When the vector $\boldsymbol{a}$ has unit norm, the dot product $\boldsymbol{a}\cdot\boldsymbol{b}$ is equal to the (signed) length of the projection of $\boldsymbol{b}$ onto $\boldsymbol{a}$. More generally,

$$\boldsymbol{a} \cdot \boldsymbol{b} = |\boldsymbol{a}|\,|\boldsymbol{b}|\,\cos\theta,$$

where θ is the angle between the two vectors, which shows that a necessary and sufficient condition for two vectors to be orthogonal is that their dot product be zero.

The *cross product* (or *outer product*) of two vectors $\boldsymbol{a} = (a_1, a_2, a_3)^T$ and $\boldsymbol{b} = (b_1, b_2, b_3)^T$ in $\mathbb{R}^3$ is the vector

$$\boldsymbol{a} \times \boldsymbol{b} \overset{\text{def}}{=} \begin{pmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{pmatrix}.$$

Note that $\boldsymbol{a} \times \boldsymbol{b} = [\boldsymbol{a}_\times]\boldsymbol{b}$, where

$$[\boldsymbol{a}_\times] \overset{\text{def}}{=} \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}.$$

The cross product of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{R}^3$ is orthogonal to these two vectors, and a necessary and sufficient condition for $\boldsymbol{a}$ and $\boldsymbol{b}$ to have the same direction is that $\boldsymbol{a} \times \boldsymbol{b} = \boldsymbol{0}$. If θ denotes as before the angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, it can be shown that

$$|\boldsymbol{a} \times \boldsymbol{b}| = |\boldsymbol{a}|\,|\boldsymbol{b}|\,|\sin\theta|.$$

TABLE 6: A syllabus for students who have a strong interest in applied mathematics, electrical engineering, or physics.

Week	Chapter	Sections	Key topics
1	1, 2	all; 2.1–2.4	cameras, shading
2	3	all	color
3	4	all	linear filters
4	5	all	building local features
5	6	all	texture
6	7	all	stereopsis
7	8	all	structure from motion with perspective cameras
8	9	all	segmentation by clustering pixels
9	10	all	fitting models
10	11	11.1–11.3	simple tracking strategies, tracking by matching, Kalman filters, data association
11	12	all	registration
12	15	all	classification
13	16	all	classifying images
14	17	all	detection
15	choice	all	one of chapters 14, 19, 20, 21

PROGRAMMING ASSIGNMENTS AND RESOURCES

The programming assignments given throughout this book sometimes require routines for numerical linear algebra, singular value decomposition, and linear and nonlinear least squares. An extensive set of such routines is available in MATLAB as well as in public-domain libraries such as LINPACK, LAPACK, and MINPACK, which can be downloaded from the Netlib repository (`http://www.netlib.org/`). In the text, we offer extensive pointers to software published on the Web and to datasets published on the Web. OpenCV is an important open-source package of computer vision routines (see Bradski and Kaehler (2008)).

ABOUT THE AUTHORS

David Forsyth received a B.Sc. (Elec. Eng.) from the University of the Witwatersrand, Johannesburg in 1984, an M.Sc. (Elec. Eng.) from that university in 1986, and a D.Phil. from Balliol College, Oxford in 1989. He spent three years on the faculty at the University of Iowa, ten years on the faculty at the University of California at Berkeley, and then moved to the University of Illinois. He served as program co-chair for IEEE Computer Vision and Pattern Recognition in 2000 and in 2011, general co-chair for CVPR 2006, and program co-chair for the European Conference on Computer Vision 2008, and is a regular member of the program committee of all major international conferences on computer vision. He has served five terms on the SIGGRAPH program committee. In 2006, he received an IEEE technical achievement award, and in 2009 he was named an IEEE Fellow.

Jean Ponce received the Doctorat de Troisieme Cycle and Doctorat d' État degrees in Computer Science from the University of Paris Orsay in 1983 and 1988. He has held Research Scientist positions at the Institut National de la Recherche en Informatique et Automatique, the MIT Artificial Intelligence Laboratory, and the Stanford University Robotics Laboratory, and served on the faculty of the Dept. of Computer Science at the University of Illinois at Urbana-Champaign from 1990 to 2005. Since 2005, he has been a Professor at Ecole Normale Superieure in Paris, France. Dr. Ponce has served on the editorial boards of Computer Vision and Image Understanding, Foundations and Trends in Computer Graphics and Vision, the IEEE Transactions on Robotics and Automation, the International Journal of Computer Vision (for which he served as Editor-in-Chief from 2003 to 2008), and the SIAM Journal on Imaging Sciences. He was Program Chair of the 1997 IEEE Conference on Computer Vision and Pattern Recognition and served as General Chair of the year 2000 edition of this conference. He also served as General Chair of the 2008 European Conference on Computer Vision. In 2003, he was named an IEEE Fellow for his contributions to Computer Vision, and he received a US patent for the development of a robotic parts feeder.