



COMPUTER VISION WITH THEME-BASED LEARNING

By Wesley Snyder

Simon J.D. Prince, *Computer Vision, Models, Learning, and Inference*, Cambridge Univ. Press, 2012, ISBN: 978-1107011793, 598 pp.



The book *Computer Vision, Models, Learning, and Inference* by Simon J.D. Prince is, unsurprisingly, about computer vision. Computer vision, in turn, is about a machine making use of information from a camera (or a camera-like device) to understand the scene the camera is viewing. Of course, this would require that we understand what we mean by *understand*. This definitional procedure could continue, but instead I'll make the same statement that Simon Prince, the author, made in the preface where he says, "Computer Vision is an engineering discipline; we're primarily motivated by the real-world concern of building machines that can see." This application-motivated approach is pursued throughout the book, even though the book is mathematically rigorous.

Most computer vision books (my own included) are technique-based. This follows our discipline's approach, which seems to value papers not based on what they do, but how they do it. As a result, there are many papers in the field with titles like "Use of the Matrix Aardvark to Find the Optimum Ant-hill." We can't, of course, go all the way to fully application-based discussions, because then the same algorithms and mathematics will be repeated ad nauseam. In this book, Prince does an outstanding job of breaking the image-understanding problem

down into the following component subproblems:

- *making measurements*—measurements could be simple things like pixel values, or more complex things like the aspect ratio of a homogeneous region;
- *building models*—a model might take a variety of forms, but in every case the model describes the relationship between the measurements and the state;
- *learning models*—these are based (usually) on parameters and these parameters must be learned (the parameters in this book are usually probabilistic in nature and frequently reflect a Bayesian approach); and
- *performing inference*—once a model has been built, the application of that model to a problem is inference (it's tempting to call this *model matching*, but it's much more than that).

After breaking the problem down into these subproblems, Prince then structures his algorithm descriptions in terms of these components.

Organization

The first three (of the six) parts of the book provide some background. Part I is about the fundamentals of probability. Part II introduces the student to density representations, regression, and classification—this part alone could be used as a limited text in

statistical pattern classification. Part III uses models for graphs, chains, trees (such as Markov), and grids to connect local information into more global models.

Part IV quickly introduces images into the problem, by talking about preprocessing. Part V is a compact restatement of much of the material in Richard Hartley and Andrew Zisserman's book, *Multiview Geometry in Computer Vision*. Part VI, "Models for Vision," delves seriously into what others have called "computer vision" by addressing shape, style, and motion.

Content

Part I (chapters 2–5) introduces the student to fundamental material on probability, including properties of the Gaussian and other common distributions and methods of estimation. Part II (chapters 6–9) covers the background material in learning. I especially like the presentation in chapter 7 on expectation maximization. Part III (chapters 10–12) is about models for images. This includes representing image information in graphical form (chapter 10), representing information by chains and trees (chapter 11), and the 2D version, representing image information on grids. The Viterbi algorithm is presented in chapter 11 and the Markov Random Field is in chapter 12. Maximum a posteriori (MAP) methods are covered in both chapters.

Part IV contains just one chapter (13), which is entitled “Preprocessing.” It starts out with whitening and histogram equalization—which I would also call preprocessing—but then it moves into linear (kernel) operators such as the Sobel, Laplacian, and Gabor. Section 13.2 moves on to edge and corner detection, including Harris and scale-invariant feature transform (or SIFT, as a detector). The histogram of oriented gradients (HOG) and SIFT (as a descriptor) are covered in section 13.3. The section on dimensionality reduction (13.4) covers principal component analysis and k-means clustering. Chapter 13 covers all these topics briefly, but clearly. In truth, this chapter covers much of the non-Bayesian aspects of computer vision.

Chapters 14–16 constitute part V, which is all about cameras, single-, stereo-, and multiple-camera vision. It includes what you would expect in this section, and does it in a manner that is easy to read and understand.

In Part VI, the book finally jumps deeply into vision by addressing the issue of “What is shape?” and represents shape using snakes, shape templates, statistical shape models, and statistical principal component analysis (PCA). This topic is carried into three dimensions. Discussion of articulated and morphable models leads to applications of understanding images of the human body. Face recognition is the motivating application behind chapter 18. Chapter 19 is titled “Temporal Models,” which refers to the models required to accomplish contour tracking, including filters such as the Kalman, extended Kalman, and particle filters. Other motion topics such as optic flow and time-to-collision aren’t addressed here. Chapter 20 points out that in computer vision, the data are really discrete and we might

think of an image or sub-image as a word to be recognized. The bag-of-words methodology is presented here. The appendix presents compact explanations of optimization (numerical, not Lagrange) and linear algebra.

Pedagogy

This is definitely a textbook and not a monograph. It’s organized as a text, with repeated themes, and excellent examples. Each chapter has a collection of appropriate homework problems to solve. Some of those problems require a computer, some don’t. The figures are outstanding. This is definitely for a graduate course—the level of mathematics, while not terribly rigorous, does require some sophistication. It would be well suited to an electrical engineering graduate student, or one in computer science or mathematics. There’s a lot of math, but that’s the nature of the field. Instead of being theoretical, however, the mathematics is problem-directed. A “notes” section is at the end of each chapter, which is an annotated biography of literature relevant to that chapter.

Many of the chapters involved with object recognition, scene recognition, and so on have separate sections specifically labeled “learning” (developing a representation) and “inference” (using the representation). This permits the student to readily identify the section’s purpose. The book endeavors to be consistent in notation, and even provides an appendix discussing that notation. The style is what I would call semiformal, in that the student is never addressed in the second person but in the imperative, and only in the homework problems. The voice is almost always passive, and only active when it provides improved readability; then the usage is “we.”

What’s Missing?

The author has done an admirable job of consolidating the discipline of computer vision into a single book, following a single theme. To accomplish this, a few topics were omitted. I miss seeing material on the Hough transform and all of its descendants—including the generalized Hough, Radon, and Simple K-Space (SKS). Another potential topic in a computer vision course is morphology. These days, morphological techniques are usually relegated to preprocessing, but that’s okay. I’m sure there are other topics that could be covered, as well. However, there aren’t enough pages in any reasonable book for everything, and these two topics don’t particularly fit the book’s probabilistic theme.

Computer Vision, Models, Learning, and Inference could be used as either a first or second course book on computer vision, depending on the instructor’s emphasis. The material is so well organized that it might be valuable as a reference after the student has completed the course. Reading this book has opened my eyes to a different teaching style. Instead of a “list of topics” approach, it’s helpful to choose a few themes and describe each topic from the perspective of that theme. It should dramatically help the student to understand the topics.

Wesley Snyder is a professor of electrical and computer engineering at North Carolina State University. His research career has been directed toward robot vision. Snyder has a PhD in electrical engineering from the University of Illinois. He’s a fellow of the IEEE and the American Institute of Medical and Biomedical Engineers. Contact him at wes@ncsu.edu.