Computer Vision

Linda Shapiro  
The University of Washington  
Seattle, Washington  
shapiro@cs.washington.edu  
and  
George Stockman  
Department of Computer Science  
Michigan State University  
East Lansing, MI  
stockman@pixel.cps.msu.edu
Contents

1 Introduction
  1.1 Machines that see.................................................. 13
  1.2 Application problems............................................. 14
  1.3 Operations on Images............................................. 15
  1.4 The Good, the Bad, and the Ugly.............................. 23
  1.5 Use of Computers and Software................................. 26
  1.6 Related Areas..................................................... 27
  1.7 The rest of the book.............................................. 28
  1.8 References....................................................... 29
  1.9 Additional Exercises.......................................... 30

2 Imaging and Image Representation.................................. 33
  2.1 Sensing Light.................................................... 33
  2.2 Imaging Devices................................................ 34
  2.3 * Problems in Digital Images.................................. 39
  2.4 Picture Functions and Digital Images......................... 41
  2.5 * Digital Image Formats....................................... 46
  2.6 Richness and Problems of Real Imagery...................... 52
  2.7 3D Structure from 2D Images ................................ 53
  2.8 Five Frames of Reference...................................... 54
  2.9 * Other Types of Sensors...................................... 56
  2.10 References.................................................... 60

3 Binary Image Analysis................................................ 63
  3.1 Pixels and Neighborhoods....................................... 63
  3.2 Applying Masks to Images..................................... 64
  3.3 Counting the Objects in an Image............................. 67
  3.4 Connected Components Labeling............................... 69
  3.5 Binary Image Morphology....................................... 75
    3.5.1 Structuring Elements..................................... 75
    3.5.2 Basic Operations......................................... 77
    3.5.3 Some Applications of Binary Morphology................ 81
    3.5.4 Conditional Dilation.................................... 85
  3.6 Region Properties............................................. 86
  3.7 Region Adjacency Graphs..................................... 95
  3.8 Thresholding Gray-Scale Images.............................. 97

3
3.8.1 The Use of Histograms for Threshold Selection .......................... 97
3.8.2 * Automatic Thresholding: the Otsu Method ............................ 99
3.9 References ............................................................................. 103

4 Pattern Recognition Concepts .................................................. 107
4.1 Pattern Recognition Problems ............................................... 107
4.2 Common model for classification ......................................... 109
4.3 Precision versus recall ....................................................... 112
4.4 Features used for representation ......................................... 113
4.5 Feature Vector Representation ............................................. 114
4.6 Implementing the Classifier ................................................ 116
4.7 Structural Techniques ......................................................... 119
4.8 The Confusion Matrix .......................................................... 122
4.9 Decision Trees .................................................................... 122
4.10 Bayesian decision-making ................................................... 128
4.11 Decisions using Multidimensional Data ............................... 132
4.12 Machines that Learn ............................................................ 135
4.13 * Artificial Neural Nets ....................................................... 136
4.14 References ............................................................................. 142

5 Filtering and Enhancing Images ................................................. 145
5.1 What needs fixing? ............................................................... 145
5.1.1 An Image needs Improvement ......................................... 146
5.1.2 Low-level features must be detected ............................... 146
5.2 Grey level mapping .............................................................. 147
5.2.1 Histogram equalization .................................................. 149
5.3 Removal of Small Image Regions ........................................ 150
5.3.1 Removal of Salt-and-Pepper Noise ................................. 151
5.3.2 Removal of Small Components ...................................... 152
5.4 Image Smoothing ............................................................... 153
5.5 Median Filtering ................................................................. 154
5.5.1 Computing an Output Image from an Input Image ............. 156
5.6 Detecting Edges using Differencing Masks ......................... 156
5.6.1 Differencing 1D Signals ............................................... 158
5.6.2 Difference Operators for 2D Images ............................... 160
5.7 Gaussian Filtering and LOG Edge Detection .................... 166
5.7.1 Detecting Edges with the LOG Filter ............................... 170
5.7.2 On Human Edge-Detection .......................................... 173
5.7.3 Marr-Hildreth Theory .................................................... 174
5.8 The Canny Edge Detector ................................................... 176
5.9 *Masks as Matched Filters .................................................. 176
5.9.1 The Vector Space of all Signals of n Samples ................. 177
5.9.2 Using an Orthogonal Basis ............................................. 179
5.9.3 Cauchy-Schwartz inequality ......................................... 181
5.9.4 The Vector Space of m x n Images ................................. 181
5.9.5 A Roberts basis for 2 x 2 neighborhoods ........................ 181
5.9.6 The Frei-Chen basis for 3x3 neighborhoods .................... 182
5.10 *Convolution and Cross Correlation ................................. 186
5.10.1 Defining operations via Masks .................................................. 186
5.10.2 The Convolution Operation ...................................................... 188
5.10.3 Possible parallel implementations .............................................. 192
5.11 *Analysis of spatial frequency using sinusoids .................................. 192
5.11.1 A Fourier Basis ........................................................................ 193
5.11.2 2D Picture Functions ................................................................. 197
5.11.3 Discrete Fourier Transform .......................................................... 199
5.11.4 Bandpass Filtering ..................................................................... 201
5.11.5 Discussion of the Fourier Transform .............................................. 203
5.11.6 *The Convolution Theorem .......................................................... 203
5.12 Summary and Discussion ................................................................ 205
5.13 References ...................................................................................... 205

6 Color and Shading ............................................................................. 209

6.1 Some Physics of Color ....................................................................... 210
6.1.1 Sensing Illuminated Objects .......................................................... 211
6.1.2 Additional Factors ....................................................................... 211
6.1.3 Sensitivity of Receptors ................................................................. 212
6.2 The RGB Basis for Color .................................................................. 213
6.3 Other Color Bases ............................................................................ 215
6.3.1 The CMY Subtractive Color System ................................................. 217
6.3.2 HSI: Hue-Saturation-Intensity ........................................................ 217
6.3.3 YIQ and YUV for TV signals .......................................................... 219
6.3.4 Using Color for Classification ....................................................... 220
6.4 Color Histograms ............................................................................ 221
6.5 Color Segmentation ......................................................................... 223
6.6 Shading ........................................................................................... 225
6.6.1 Radiation from One Light Source .................................................... 225
6.6.2 Diffuse Reflection ........................................................................ 226
6.6.3 Specular reflection ....................................................................... 228
6.6.4 Darkening with Distance ............................................................... 229
6.6.5 Complications ............................................................................. 230
6.6.6 * Phong Model of Shading ............................................................. 231
6.6.7 Human Perception using Shading ................................................... 231
6.7 * Related Topics .............................................................................. 231
6.7.1 Applications ............................................................................... 231
6.7.2 Human Color Perception ............................................................... 232
6.7.3 Multispectral Images .................................................................. 232
6.7.4 Thematic Images ......................................................................... 233
6.8 References ...................................................................................... 233

7 Texture ............................................................................................... 235

7.1 Texture, Texels, and Statistics .......................................................... 236
7.2 Texel-Based Texture Descriptions .................................................... 237
7.3 Quantitative Texture Measures .......................................................... 238
7.3.1 Edge Density and Direction .......................................................... 239
7.3.2 Local Binary Partition ................................................................. 240
7.3.3 Co-occurrence Matrices and Features ........................................... 240
7.3.4 Laws’ Texture Energy Measures ........................................ 243
7.3.5 Autocorrelation and Power Spectrum ......................... 244
7.4 Texture Segmentation .................................................. 245
7.5 References ................................................................ 247

8 Content-Based Image Retrieval ........................................ 249
8.1 Image Database Examples .............................................. 249
8.2 Image Database Queries ............................................... 251
8.3 Query-by-Example ....................................................... 252
8.4 Image Distance Measures .............................................. 254
  8.4.1 Color Similarity Measures ........................................ 254
  8.4.2 Texture Similarity Measures ....................................... 257
  8.4.3 Shape Similarity Measures ......................................... 259
  8.4.4 Object Presence and Relational Similarity Measures ...... 264
8.5 Database Organization .................................................. 268
  8.5.1 Standard Indexes ..................................................... 268
  8.5.2 Spatial Indexing ...................................................... 271
  8.5.3 Indexing for Content-Based Image Retrieval with Multiple Distance Measures ........................................ 271
8.6 References ................................................................ 273

9 Motion from 2D Image Sequences ...................................... 275
9.1 Motion Phenomena and Applications ................................ 275
9.2 Image Subtraction ........................................................ 277
9.3 Computing Motion Vectors ............................................. 279
  9.3.1 The Decathlon Game ................................................ 280
  9.3.2 Using Point Correspondences ................................... 281
  9.3.3 MPEG C Compression of Video ............................... 285
  9.3.4 * Computing Image Flow .......................................... 287
  9.3.5 * The Image Flow Equation ...................................... 288
  9.3.6 * Solving for Image Flow by Propagating Constraints .... 289
9.4 Computing the Paths of Moving Points ......................... 290
  9.4.1 Integrated Problem-Specific Tracking ....................... 296
9.5 Detecting Significant Changes in Video ......................... 298
  9.5.1 Segmenting Video Sequences ................................... 299
  9.5.2 Ignoring Certain Camera Effects ............................... 301
  9.5.3 Storing Video Subsequences .................................... 302
9.6 References ................................................................ 303

10 Image Segmentation ........................................................ 305
10.1 Identifying Regions ..................................................... 306
  10.1.1 Clustering Methods ................................................. 307
  10.1.2 Region Growing ..................................................... 315
10.2 Representing Regions .................................................... 318
  10.2.1 Overlays ............................................................... 318
  10.2.2 Labeled Images ...................................................... 318
  10.2.3 Boundary Coding ................................................... 320
  10.2.4 Quad Trees ........................................................... 320
10.2.5 Property Tables .................................................. 322
10.3 Identifying Contours ................................................. 322
10.3.1 Tracking Existing Region Boundaries ....................... 323
10.3.2 The Canny Edge Detector and Linker ...................... 326
10.3.3 Aggregating Consistent Neighboring Edges into Curves . 329
10.3.4 Hough Transform for Lines and Circular Arcs .......... 330
10.4 Fitting Models to Segments ....................................... 341
10.5 Identifying Higher-level Structure .............................. 346
10.5.1 Ribbons .......................................................... 346
10.5.2 Detecting Corners .............................................. 348
10.6 Segmentation using Motion Coherence ....................... 349
10.6.1 Boundaries in Space-Time ................................. 350
10.6.2 Aggregating Motion Trajectories ......................... 350
10.7 References .......................................................... 351

11 Matching in 2D ....................................................... 357
11.1 Registration of 2D Data ............................................. 357
11.2 Representation of Points ......................................... 359
11.3 Affine Mapping Functions ...................................... 360
11.4 * A Best 2D Affine Transformation ........................... 371
11.5 2D Object Recognition via Affine Mapping ................. 372
11.6 2D Object Recognition via Relational Matching .......... 382
11.7 Nonlinear Warping ................................................ 398
11.8 Summary .......................................................... 402
11.9 References .......................................................... 402

12 Perceiving 3D from 2D Images .................................. 405
12.1 Intrinsic Images .................................................... 405
12.2 Labeling of Line Drawings from Blocks World .......... 410
12.3 3D Cues Available in 2D Images .............................. 417
12.4 Other Phenomena .................................................. 422
  12.4.1 Shape from X ................................................ 422
  12.4.2 Vanishing Points .......................................... 426
  12.4.3 Depth from Focus ......................................... 427
  12.4.4 Motion Phenomena ....................................... 428
  12.4.5 Boundaries and Virtual Lines ......................... 428
  12.4.6 Alignments are Non-Accidental ....................... 428
12.5 The Perspective Imaging Model ................................. 429
12.6 Depth Perception from Stereo ................................. 431
  12.6.1 Establishing Correspondences .......................... 434
12.7 * The Thin Lens Equation ..................................... 438
12.8 Concluding Discussion .......................................... 441
12.9 References .......................................................... 442
## 13 3D Sensing and Object Pose Computation

13.1 General Stereo Configuration .............................................. 446
13.2 3D Affine Transformations .................................................. 448
  13.2.1 Coordinate Frames ..................................................... 448
  13.2.2 Translation ............................................................... 450
  13.2.3 Scaling ................................................................. 450
  13.2.4 Rotation ................................................................. 450
  13.2.5 Arbitrary Rotation ..................................................... 453
  13.2.6 Alignment via Transformation Calculus ............................ 454
13.3 Camera Model ................................................................. 454
  13.3.1 Perspective Transformation Matrix ................................ 458
  13.3.2 Orthographic and Weak Perspective Projections ................ 461
  13.3.3 Computing 3D Points Using Multiple Cameras .................. 463
13.4 Best Affine Calibration Matrix ............................................ 465
  13.4.1 Calibration Jig .......................................................... 466
  13.4.2 Defining the Least-Squares Problem ............................... 466
  13.4.3 Discussion of the Affine Method .................................... 471
13.5 Using Structured Light ...................................................... 472
13.6 A Simple Pose Estimation Procedure ................................... 474
13.7 * An Improved Camera Calibration Method ............................. 479
  13.7.1 Intrinsic Camera Parameters ....................................... 480
  13.7.2 Extrinsic Camera Parameters ...................................... 481
  13.7.3 Calibration Example .................................................. 485
13.8 * Pose Estimation ............................................................ 489
  13.8.1 Pose from 2D-3D Point Correspondences ......................... 491
  13.8.2 Constrained Linear Optimization .................................. 492
  13.8.3 Computing the Transformation $\mathbf{T} = \{\mathbf{R}, \mathbf{T}\} $ ....... 493
  13.8.4 Verification and Optimization of Pose ............................ 495
13.9 3D Object Reconstruction .................................................. 496
  13.9.1 Data Acquisition ..................................................... 497
  13.9.2 Registration of Views ................................................ 499
  13.9.3 Surface Reconstruction ............................................. 500
  13.9.4 Space-Carving ....................................................... 500
13.10 Computing Shape from Shading ......................................... 505
  13.10.1 Photometric Stereo ................................................ 508
  13.10.2 Integrating Spatial Constraints .................................... 509
13.11 Structure from Motion .................................................... 511
13.12 References ................................................................. 514

## 14 3D Models and Matching

14.1 Survey of Common Representation Methods ............................ 518
  14.1.1 3D Mesh Models ..................................................... 518
  14.1.2 Surface-Edge-Vertex Models ...................................... 518
  14.1.3 Generalized-Cylinder Models ..................................... 521
  14.1.4 Octrees ............................................................... 524
  14.1.5 Superquadrics ....................................................... 525
14.2 True 3D Models versus View-Class Models ............................. 526
14.3 Physics-based and Deformable Models .................................. 528
14.3.1 Snakes: Active Contour Models ........................................... 528
14.3.2 Balloon Models for 3D ....................................................... 531
14.3.3 Modeling Motion of the Human Heart ............................. 532
14.4 3D Object Recognition Paradigms ........................................ 532
14.4.1 Matching Geometric Models via Alignment ..................... 536
14.4.2 Matching Relational Models .............................................. 542
14.4.3 Matching Functional Models ............................................ 554
14.4.4 Recognition by Appearance ............................................. 560
14.5 References ................................................................. 567

15 Virtual Reality ........................................................................ 571
15.1 Features of Virtual Reality Systems ...................................... 572
15.2 Applications of VR ............................................................ 572
15.3 Augmented Reality (AR) ...................................................... 574
15.4 Teleoperation ................................................................. 576
15.5 Virtual Reality Devices ...................................................... 580
15.6 Summary of Sensing Devices for VR ................................. 585
15.7 Rendering Simple 3D Models .............................................. 587
15.8 Composing Real and Synthetic Imagery ......................... 589
15.9 HCI and Psychological Issues ............................................. 593
15.10 References ................................................................. 593

16 Case Studies ........................................................................ 595
16.1.1 Application Domain and Requirements ......................... 597
16.1.2 System Design .......................................................... 597
16.1.3 Identification Procedure .................................................. 598
16.1.4 More Details on the Process .......................................... 598
16.1.5 Performance ............................................................ 601
16.2 Identifying Humans via the Iris of an Eye ......................... 602
16.2.1 Requirements for identification systems ......................... 602
16.2.2 System Design .......................................................... 604
16.2.3 Performance ............................................................ 607
16.2.4 References ............................................................ 609
Preface

This book is intended as an introduction to computer vision for a broad audience. It provides necessary theory and examples for students and practitioners who will work in fields where significant information must be extracted automatically from images. The book should be a useful resource book for professionals, a text for both undergraduate and beginning graduate courses, and a resource for enrichment of college or even high school projects. Our goals were to provide a basic set of fundamental concepts and algorithms and also discuss some of the exciting evolving application areas. This book is unique in that it contains chapters on image databases (Ch 8) and on virtual and augmented reality (Ch 15), two exciting evolving application areas. A final chapter (Ch 16) gives a complete view of real world systems that use computer vision.

Due to recent progress in the computer field, economical and flexible use of computer images is now pervasive. Computing with images is no longer just for the realm of the sciences, but also for the arts and social sciences and even for hobbyists. The book should serve an established and growing audience including those interested in multimedia, art and design, geographic information systems, and image databases, in addition to the traditional areas of automation, image science, medical imaging, remote sensing and computer cartography.

A broad purpose at first seems impossible to achieve. However, there are other kinds of texts that already do this in other areas – calculus, physics, and general computing. We hope we have made at least a good beginning – we wanted a book that would be useful in the classroom and also to the independent reader. We find the chosen topics interesting and sometimes exciting, and hope that they are accessible to a large audience. It is assumed that use of the text in a graduate, or even senior level, computer vision course would be supplemented by papers from the archival literature. Coverage is not intended to be comprehensive; only a modest set of papers are cited at the end of each chapter.

The early chapters begin at an intuitive level and progress towards mathematical models with the goal of intuitive understanding before formal characterization. Sections marked by "*" are more mathematical or more advanced and need not be covered in a less technical course. To strengthen the intuitive approach, we have stayed with the processing of iconic imagery for the first eleven chapters and have delayed 3D computer vision until the later chapters, but it should be easy for experienced instructors to resequence them to fit a particular course or teaching style. There are many viable applications that are entirely 2D, and many concepts and algorithms are more simply taught in their 2D form. We provide some basics of pattern recognition in Chapter 4, so that students can consider complete recognition systems before the full coverage of image features and matching. A reader should have a good idea of 2D image processing applications after Chapter 4; Chapters 5, 6, and 7 add in gray-tone, color, and texture features. Chapter 8 treats image databases, a popular recent topic. Although some colleagues advised us to place this material near the end of the book, our goal of positioning it early in the chapter sequence is to reinforce the concepts of the prior chapters and to provide material that can lead to an excellent half-term project. Segmentation and matching are treated in their 2D forms in Chapters 10 and 11, so that the basic concepts are presented in a simple form, without introducing the complexities of 3D transformations.
Characteristics of the 3D world are briefly introduced in Chapter 2 and then are studied in much more detail in Chapter 12. Chapter 12 surveys qualitatively many aspects of how a 3D world can be perceived from 2D images: it concludes with quantitative models of stereo and study of the thin lens equation for depth-from-focus and resolving power. The transition to 3D computer vision is made in Chapter 13: the authors have found from their own teaching that the difficulty increases abruptly for students at this point. The use of matrices to model homogeneous transformations are included within the chapter rather than in appendices; the 3D versions are extensions of the simpler 2D versions given in Chapter 11. Least-squares fitting, introduced in a simple 2D context in Chapter 11, is also extended in Chapter 13. Non-linear optimization is introduced in a simple P3P context and then used for camera calibration including the modeling of radial distortion in a lens. Chapter 14 treats 3D models and the matching of models to 3D sensed data: it is of mixed difficulty. Chapter 15 discusses applications in virtual and augmented (mixed) reality and the role of computer vision techniques.

Programming Language Issue

The book does not rely on any programming language, but uses a generic algorithmic notation. Commitment to a particular language is unnecessary and would be the wrong language for many readers. Students who are programmers should have little trouble implementing the algorithms, as our own students have shown. Examples will eventually be provided on the WWW when appropriate and available, primarily so students can quickly experiment, secondarily so that they can study some sample code.

Several tools and libraries are available to instructors and students; for example, Khoros, NIH-Image, XView, gimp, MATLAB, etc. There are also packages that can be purchased from companies that make machine vision hardware. The authors have decided not to base the text on any specific software because, first, most readers would be using something else, and second, it would be counterproductive to bury the essence of the image operations within the complex framework of data structures and methods needed in an industrial strength system. Having first studied principles in an environment with few variables, the reader will then be better able to successfully choose and use an industrial system.

Ways to use the text

The book material can be selected, and sometimes sequenced, in different ways according to the goal of the course and interests of the instructor and students.

- Chapter 3, with brief summary of Chapter 2
  A minimum usage would be 1-3 lectures in a data structures and algorithms course. Chapter 3, with some background from Chapter 2 contains motivational applications and programming exercises on 2D arrays, depth-first search, and the UNION-FIND operation for sets.

- Chapters 1, 2, 3, and optionally some of [4, 5, 6]
  This could serve as an enrichment unit of 1 to 3 weeks for high school or lower division undergrads. The objective could be as simple as a term paper or as complex as group
work on a program to, say, create a 2D parts recognition system based on connected components and prototype matching of feature vectors.

- Much of Chapters 1-11
  This would be a survey of 2D material for an elective course for students in geography, natural resources or microbiology, for example, provided many of the optional sections are passed over. If most sections of Chapters 1-11 are covered, this would constitute a semester undergraduate course in image processing and analysis with an introduction to computer vision.

- Most of the text
  This would constitute a semester course in computer vision for the senior or first year graduate student level. There is more material in the book than can be covered well in one semester. Some sections will have to be ignored or surveyed and the reader should not be expected to be able to work homework problems in all sections. For the quarter system, Chapters 1-4, 6-12, and 14 make a good introduction to computer vision for undergraduates. For a one quarter graduate course, Chapters 1-4 can be minimally covered with the emphasis on Chapters 6-14 and a brief coverage of chapter 15. For any graduate level course, it is expected that some papers from the current literature would also be covered.

We’re grateful to our many colleagues, teachers, and students with whom we have shared our interests. They have contributed much to our growing field and shared their work and excitement. Many have generously supported this book with encouragement and with contributions of ideas, figures, and algorithms. Specific citations are given throughout the book. With regret we have left out some important contributions — a text can only be so large. The several reviewers and many colleagues who have given us feedback have significantly improved our work. In particular, for careful editing, we are indebted to Mohammad Ghavamzadeh, Nick Dutta, Kevin Bowyer, Adam Clark, Yu-Yu Chou, Habib Abi-Rached, and Valentin Razmov. We take responsibility for any errors remaining in the book and for providing corrections in the future.

Linda Shapiro and George Stockman (March 2000)
(shapiro@cs.washington.edu and stockman@cse.msu.edu)