

INTERNATIONAL EDITION

Computer Vision

A MODERN APPROACH

FORSYTH ■ PONCE



Contents

Part I Image Formation and Image Models	1
1 CAMERAS	3
1.1 Pinhole Cameras 4	
1.1.1 Perspective Projection, 4	
1.1.2 Affine Projection, 6	
1.2 Cameras with Lenses 7	
1.2.1 Paraxial Geometric Optics, 8	
1.2.2 Thin Lenses, 9	
1.2.3 Real Lenses, 11	
1.3 The Human Eye 13	
1.4 Sensing 15	
1.4.1 CCD Cameras, 16	
1.4.2 Sensor Models, 17	
1.5 Notes 18	
Problems 19	
2 GEOMETRIC CAMERA MODELS	20
2.1 Elements of Analytical Euclidean Geometry 20	
2.1.1 Coordinate Systems and Homogeneous Coordinates, 20	
2.1.2 Coordinate System Changes and Rigid Transformations, 23	
2.2 Camera Parameters and the Perspective Projection 28	
2.2.1 Intrinsic Parameters, 29	

- 2.2.2 Extrinsic Parameters, 30
- 2.2.3 A Characterization of Perspective Projection Matrices, 31
- 2.3 Affine Cameras and Affine Projection Equations 32
 - 2.3.1 Affine Cameras, 32
 - 2.3.2 Affine Projection Equations, 33
 - 2.3.3 A Characterization of Affine Projection Matrices, 34
- 2.4 Notes 35
 - Problems 35

3 GEOMETRIC CAMERA CALIBRATION

38

- 3.1 Least-Squares Parameter Estimation 39
 - 3.1.1 Linear Least-Squares Methods, 39
 - 3.1.2 Nonlinear Least-Squares Methods, 42
- 3.2 A Linear Approach to Camera Calibration 45
 - 3.2.1 Estimation of the Projection Matrix, 45
 - 3.2.2 Estimation of the Intrinsic and Extrinsic Parameters, 45
 - 3.2.3 Degenerate Point Configurations, 46
- 3.3 Taking Radial Distortion into Account 47
 - 3.3.1 Estimation of the Projection Matrix, 48
 - 3.3.2 Estimation of the Intrinsic and Extrinsic Parameters, 48
 - 3.3.3 Degenerate Point Configurations, 49
- 3.4 Analytical Photogrammetry 50
- 3.5 An Application: Mobile Robot Localization 51
- 3.6 Notes 53
 - Problems 53

4 RADIOMETRY—MEASURING LIGHT

55

- 4.1 Light in Space 55
 - 4.1.1 Foreshortening, 55
 - 4.1.2 Solid Angle, 56
 - 4.1.3 Radiance, 58
- 4.2 Light at Surfaces 60
 - 4.2.1 Simplifying Assumptions, 60
 - 4.2.2 The Bidirectional Reflectance Distribution Function, 60
 - 4.2.3 Example: The Radiometry of Thin Lenses, 62
- 4.3 Important Special Cases 63
 - 4.3.1 Radiosity, 63
 - 4.3.2 Directional Hemispheric Reflectance, 64
 - 4.3.3 Lambertian Surfaces and Albedo, 65
 - 4.3.4 Specular Surfaces, 65
 - 4.3.5 The Lambertian + Specular Model, 66
- 4.4 Notes 67
 - Problems 68

5 SOURCES, SHADOWS, AND SHADING

70

- 5.1 Qualitative Radiometry 70
- 5.2 Sources and Their Effects 71
 - 5.2.1 Radiometric Properties of Light Sources, 71
 - 5.2.2 Point Sources, 72

- 5.2.3 Line Sources, 75
- 5.2.4 Area Sources, 75
- 5.3 Local Shading Models 77
 - 5.3.1 Local Shading Models for Point Sources, 77
 - 5.3.2 Area Sources and Their Shadows, 79
 - 5.3.3 Ambient Illumination, 79
- 5.4 Application: Photometric Stereo 80
 - 5.4.1 Normal and Albedo from Many Views, 82
 - 5.4.2 Shape from Normals, 84
- 5.5 Interreflections: Global Shading Models 86
 - 5.5.1 An Interreflection Model, 87
 - 5.5.2 Solving for Radiosity, 89
 - 5.5.3 The Qualitative Effects of Interreflections, 90
- 5.6 Notes 91
 - Problems 95

6 COLOR

97

- 6.1 The Physics of Color 97
 - 6.1.1 Radiometry for Colored Lights: Spectral Quantities, 97
 - 6.1.2 The Color of Sources, 98
 - 6.1.3 The Color of Surfaces, 100
- 6.2 Human Color Perception 102
 - 6.2.1 Color Matching, 102
 - 6.2.2 Color Receptors, 104
- 6.3 Representing Color 105
 - 6.3.1 Linear Color Spaces, 105
 - 6.3.2 Non-linear Color Spaces, 110
 - 6.3.3 Spatial and Temporal Effects, 113
- 6.4 A Model for Image Color 114
 - 6.4.1 Cameras, 114
 - 6.4.2 A Model for Image Color, 115
 - 6.4.3 Application: Finding Specularities, 118
- 6.5 Surface Color from Image Color 119
 - 6.5.1 Surface Color Perception in People, 120
 - 6.5.2 Inferring Lightness, 122
 - 6.5.3 Surface Color from Finite-Dimensional Linear Models, 126
- 6.6 Notes 128
 - Problems 130

Part II Early Vision: Just One Image

133

7 LINEAR FILTERS

135

- 7.1 Linear Filters and Convolution 135
 - 7.1.1 Convolution, 136
- 7.2 Shift Invariant Linear Systems 140
 - 7.2.1 Discrete Convolution, 141
 - 7.2.2 Continuous Convolution, 142
 - 7.2.3 Edge Effects in Discrete Convolutions, 145

- 7.3 Spatial Frequency and Fourier Transforms 145
 - 7.3.1 Fourier Transforms, 146
- 7.4 Sampling and Aliasing 148
 - 7.4.1 Sampling, 150
 - 7.4.2 Aliasing, 152
 - 7.4.3 Smoothing and Resampling, 153
- 7.5 Filters as Templates 157
 - 7.5.1 Convolution as a Dot Product, 157
 - 7.5.2 Changing Basis, 158
- 7.6 Technique: Normalized Correlation and Finding Patterns 158
 - 7.6.1 Controlling the Television by Finding Hands by Normalized Correlation, 158
- 7.7 Technique: Scale and Image Pyramids 159
 - 7.7.1 The Gaussian Pyramid, 159
 - 7.7.2 Applications of Scaled Representations, 161
- 7.8 Notes 162
 - Problems 164

8 EDGE DETECTION

165

- 8.1 Noise 165
 - 8.1.1 Additive Stationary Gaussian Noise, 166
 - 8.1.2 Why Finite Differences Respond to Noise, 168
- 8.2 Estimating Derivatives 169
 - 8.2.1 Derivative of Gaussian Filters, 169
 - 8.2.2 Why Smoothing Helps, 170
 - 8.2.3 Choosing a Smoothing Filter, 171
 - 8.2.4 Why Smooth with a Gaussian? 172
- 8.3 Detecting Edges 175
 - 8.3.1 Using the Laplacian to Detect Edges, 175
 - 8.3.2 Gradient-Based Edge Detectors, 176
 - 8.3.3 Technique: Orientation Representations and Corners, 181
- 8.4 Notes 185
 - Problems 187

9 TEXTURE

189

- 9.1 Representing Texture 190
 - 9.1.1 Extracting Image Structure with Filter Banks, 191
 - 9.1.2 Representing Texture Using the Statistics of Filter Outputs, 193
- 9.2 Analysis (and Synthesis) Using Oriented Pyramids 196
 - 9.2.1 The Laplacian Pyramid, 197
 - 9.2.2 Filters in the Spatial Frequency Domain, 199
 - 9.2.3 Oriented Pyramids, 202
- 9.3 Application: Synthesizing Textures for Rendering 202
 - 9.3.1 Homogeneity, 205
 - 9.3.2 Synthesis by Sampling Local Models, 205
- 9.4 Shape from Texture 207
 - 9.4.1 Shape from Texture for Planes, 208
- 9.5 Notes 210
 - Problems 211

Part III Early Vision: Multiple Images	213
10 THE GEOMETRY OF MULTIPLE VIEWS	215
10.1 Two Views 216	
10.1.1 Epipolar Geometry, 216	
10.1.2 The Calibrated Case, 217	
10.1.3 Small Motions, 218	
10.1.4 The Uncalibrated Case, 218	
10.1.5 Weak Calibration, 219	
10.2 Three Views 222	
10.2.1 Trifocal Geometry, 223	
10.2.2 The Calibrated Case, 223	
10.2.3 The Uncalibrated Case, 226	
10.2.4 Estimation of the Trifocal Tensor, 226	
10.3 More Views 227	
10.4 Notes 230	
Problems 232	
11 STEREOPSIS	234
11.1 Reconstruction 235	
11.1.1 Image Rectification, 236	
11.2 Human Stereopsis 237	
11.3 Binocular Fusion 240	
11.3.1 Correlation, 240	
11.3.2 Multi-Scale Edge Matching, 242	
11.3.3 Dynamic Programming, 243	
11.4 Using More Cameras 246	
11.4.1 Three Cameras, 246	
11.4.2 Multiple Cameras, 247	
11.5 Notes 247	
Problems 250	
12 AFFINE STRUCTURE FROM MOTION	251
12.1 Elements of Affine Geometry 253	
12.1.1 Affine Spaces and Barycentric Combinations, 253	
12.1.2 Affine Subspaces and Affine Coordinates, 254	
12.1.3 Affine Transformations and Affine Projection Models, 257	
12.1.4 Affine Shape, 258	
12.2 Affine Structure and Motion from Two Images 258	
12.2.1 Geometric Scene Reconstruction, 259	
12.2.2 Algebraic Motion Estimation, 260	
12.3 Affine Structure and Motion from Multiple Images 263	
12.3.1 The Affine Structure of Affine Image Sequences, 263	
12.3.2 A Factorization Approach to Affine Structure from Motion, 263	
12.4 From Affine to Euclidean Images 266	
12.4.1 Euclidean Constraints and Calibrated Affine Cameras, 266	
12.4.2 Computing Euclidean Upgrades from Multiple Views, 267	
12.5 Affine Motion Segmentation 269	
12.5.1 The Reduced Row-Echelon Form of the Data Matrix, 269	

- 12.5.2 The Shape Interaction Matrix, 270
- 12.6 Notes 272
- Problems 272

13 PROJECTIVE STRUCTURE FROM MOTION 274

- 13.1 Elements of Projective Geometry 275
 - 13.1.1 Projective Spaces, 275
 - 13.1.2 Projective Subspaces and Projective Coordinates, 276
 - 13.1.3 Affine and Projective Spaces, 278
 - 13.1.4 Hyperplanes and Duality, 279
 - 13.1.5 Cross-Ratios and Projective Coordinates, 280
 - 13.1.6 Projective Transformations, 282
 - 13.1.7 Projective Shape, 284
- 13.2 Projective Structure and Motion from Binocular Correspondences 284
 - 13.2.1 Geometric Scene Reconstruction, 284
 - 13.2.2 Algebraic Motion Estimation, 285
- 13.3 Projective Motion Estimation from Multilinear Constraints 287
 - 13.3.1 Motion Estimation from Fundamental Matrices, 288
 - 13.3.2 Motion Estimation from Trifocal Tensors, 288
- 13.4 Projective Structure and Motion from Multiple Images 289
 - 13.4.1 A Factorization Approach to Projective Structure from Motion, 289
 - 13.4.2 Bundle Adjustment, 292
- 13.5 From Projective to Euclidean Images 292
- 13.6 Notes 294
- Problems 296

Part IV Mid-Level Vision 299

14 SEGMENTATION BY CLUSTERING 301

- 14.1 What Is Segmentation? 301
 - 14.1.1 Model Problems, 303
 - 14.1.2 Segmentation as Clustering, 304
- 14.2 Human Vision: Grouping and Gestalt 304
- 14.3 Applications: Shot Boundary Detection and Background Subtraction 309
 - 14.3.1 Background Subtraction, 309
 - 14.3.2 Shot Boundary Detection, 310
- 14.4 Image Segmentation by Clustering Pixels 313
 - 14.4.1 Segmentation Using Simple Clustering Methods, 313
 - 14.4.2 Clustering and Segmentation by K-means, 315
- 14.5 Segmentation by Graph-Theoretic Clustering 317
 - 14.5.1 Terminology for Graphs, 317
 - 14.5.2 The Overall Approach, 318
 - 14.5.3 Affinity Measures, 319
 - 14.5.4 Eigenvectors and Segmentation, 321
 - 14.5.5 Normalized Cuts, 323
- 14.6 Notes 326
- Problems 327

15	SEGMENTATION BY FITTING A MODEL	329
15.1	The Hough Transform 329	
15.1.1	Fitting Lines with the Hough Transform, 330	
15.1.2	Practical Problems with the Hough Transform, 330	
15.2	Fitting Lines 333	
15.2.1	Line Fitting with Least Squares, 333	
15.2.2	Which Point Is on Which Line? 335	
15.3	Fitting Curves 337	
15.3.1	Implicit Curves, 337	
15.3.2	Parametric Curves, 340	
15.4	Fitting as a Probabilistic Inference Problem 341	
15.5	Robustness 342	
15.5.1	M-estimators, 343	
15.5.2	RANSAC, 346	
15.6	Example: Using RANSAC to Fit Fundamental Matrices 348	
15.6.1	An Expression for Fitting Error, 348	
15.6.2	Correspondence as Noise, 348	
15.6.3	Applying RANSAC, 348	
15.6.4	Finding the Distance, 349	
15.6.5	Fitting a Fundamental Matrix to Known Correspondences, 351	
15.7	Notes 351	
	Problems 352	
16	SEGMENTATION AND FITTING USING PROBABILISTIC METHODS	354
16.1	Missing Data Problems, Fitting, and Segmentation 354	
16.1.1	Missing Data Problems, 355	
16.1.2	The EM Algorithm, 357	
16.1.3	The EM Algorithm in the General Case, 359	
16.2	The EM Algorithm in Practice 359	
16.2.1	Example: Image Segmentation, Revisited, 359	
16.2.2	Example: Line Fitting with EM, 362	
16.2.3	Example: Motion Segmentation and EM, 363	
16.2.4	Example: Using EM to Identify Outliers, 365	
16.2.5	Example: Background Subtraction Using EM, 367	
16.2.6	Example: EM and the Fundamental Matrix, 368	
16.2.7	Difficulties with the EM Algorithm, 368	
16.3	Model Selection: Which Model Is the Best Fit? 369	
16.3.1	Basic Ideas, 369	
16.3.2	AIC—An Information Criterion, 369	
16.3.3	Bayesian Methods and Schwartz' BIC, 370	
16.3.4	Description Length, 370	
16.3.5	Other Methods for Estimating Deviance, 370	
16.4	Notes 371	
	Problems 372	
17	TRACKING WITH LINEAR DYNAMIC MODELS	373
17.1	Tracking as an Abstract Inference Problem 374	
17.1.1	Independence Assumptions, 374	

- 17.1.2 Tracking as Inference, 375
- 17.1.3 Overview, 376
- 17.2 Linear Dynamic Models 376
 - 17.2.1 Drifting Points, 377
 - 17.2.2 Constant Velocity, 377
 - 17.2.3 Constant Acceleration, 377
 - 17.2.4 Periodic Motion, 379
 - 17.2.5 Higher Order Models, 379
- 17.3 Kalman Filtering 380
 - 17.3.1 The Kalman Filter for a 1D State Vector, 380
 - 17.3.2 The Kalman Update Equations for a General State Vector, 383
 - 17.3.3 Forward–Backward Smoothing, 383
- 17.4 Data Association 388
 - 17.4.1 Choosing the Nearest—Global Nearest Neighbours, 389
 - 17.4.2 Gating and Probabilistic Data Association, 390
- 17.5 Applications and Examples 393
 - 17.5.1 Vehicle Tracking, 393
- 17.6 Notes 397
 - Problems 397

Part V High-Level Vision: Geometric Methods 399

18 MODEL-BASED VISION 401

- 18.1 Initial Assumptions 401
 - 18.1.1 Obtaining Hypotheses, 402
- 18.2 Obtaining Hypotheses by Pose Consistency 403
 - 18.2.1 Pose Consistency for Perspective Cameras, 403
 - 18.2.2 Affine and Projective Camera Models, 404
 - 18.2.3 Linear Combinations of Models, 406
- 18.3 Obtaining Hypotheses by Pose Clustering 407
- 18.4 Obtaining Hypotheses Using Invariants 410
 - 18.4.1 Invariants for Plane Figures, 410
 - 18.4.2 Geometric Hashing, 413
 - 18.4.3 Invariants and Indexing, 413
- 18.5 Verification 416
 - 18.5.1 Edge Proximity, 416
 - 18.5.2 Similarity in Texture, Pattern, and Intensity, 418
- 18.6 Application: Registration in Medical Imaging Systems 418
 - 18.6.1 Imaging Modes, 419
 - 18.6.2 Applications of Registration, 420
 - 18.6.3 Geometric Hashing Techniques in Medical Imaging, 421
- 18.7 Curved Surfaces and Alignment 422
- 18.8 Notes 423
 - Problems 425

19 SMOOTH SURFACES AND THEIR OUTLINES 427

- 19.1 Elements of Differential Geometry 429
 - 19.1.1 Curves, 429
 - 19.1.2 Surfaces, 434

- 19.2 Contour Geometry 438
 - 19.2.1 The Occluding Contour and the Image Contour, 439
 - 19.2.2 The Cusps and Inflections of the Image Contour, 439
 - 19.2.3 Koenderink's Theorem, 441
- 19.3 Notes 442
 - Problems 443

20 ASPECT GRAPHS

444

- 20.1 Visual Events: More Differential Geometry 447
 - 20.1.1 The Geometry of the Gauss Map, 448
 - 20.1.2 Asymptotic Curves, 449
 - 20.1.3 The Asymptotic Spherical Map, 450
 - 20.1.4 Local Visual Events, 451
 - 20.1.5 The Bitangent Ray Manifold, 453
 - 20.1.6 Multilocal Visual Events, 455
- 20.2 Computing the Aspect Graph 456
 - 20.2.1 Step 1: Tracing Visual Events, 457
 - 20.2.2 Step 2: Constructing the Regions, 459
 - 20.2.3 Remaining Steps of the Algorithm, 459
 - 20.2.4 An Example, 460
- 20.3 Aspect Graphs and Object Localization 460
- 20.4 Notes 464
 - Problems 465

21 RANGE DATA

467

- 21.1 Active Range Sensors 467
- 21.2 Range Data Segmentation 469
 - 21.2.1 Elements of Analytical Differential Geometry, 469
 - 21.2.2 Finding Step and Roof Edges in Range Images, 471
 - 21.2.3 Segmenting Range Images into Planar Regions, 476
- 21.3 Range Image Registration and Model Acquisition 477
 - 21.3.1 Quaternions, 478
 - 21.3.2 Registering Range Images Using the Iterative Closest-Point Method, 479
 - 21.3.3 Fusing Multiple Range Images, 481
- 21.4 Object Recognition 483
 - 21.4.1 Matching Piecewise-Planar Surfaces Using Interpretation Trees, 483
 - 21.4.2 Matching Free-Form Surfaces Using Spin Images, 486
- 21.5 Notes 490
 - Problems 491

Part VI High-Level Vision: Probabilistic and Inferential Methods

493

22 FINDING TEMPLATES USING CLASSIFIERS

495

- 22.1 Classifiers 496
 - 22.1.1 Using Loss to Determine Decisions, 496
 - 22.1.2 Overview: Methods for Building Classifiers, 497

- 22.1.3 Example: A Plug-in Classifier for Normal Class-conditional Densities, 499
- 22.1.4 Example: A Nonparametric Classifier Using Nearest Neighbors, 500
- 22.1.5 Estimating and Improving Performance, 501
- 22.2 Building Classifiers from Class Histograms 502
 - 22.2.1 Finding Skin Pixels Using a Classifier, 503
 - 22.2.2 Face Finding Assuming Independent Template Responses, 504
- 22.3 Feature Selection 505
 - 22.3.1 Principal Component Analysis, 507
 - 22.3.2 Identifying Individuals with Principal Components Analysis, 508
 - 22.3.3 Canonical Variates, 512
- 22.4 Neural Networks 516
 - 22.4.1 Key Ideas, 516
 - 22.4.2 Minimizing the Error, 518
 - 22.4.3 When to Stop Training, 520
 - 22.4.4 Finding Faces Using Neural Networks, 520
 - 22.4.5 Convolutional Neural Nets, 522
- 22.5 The Support Vector Machine 523
 - 22.5.1 Support Vector Machines for Linearly Separable Datasets, 524
 - 22.5.2 Finding Pedestrians Using Support Vector Machines, 527
- 22.6 Notes 528
 - Problems 531
- 22.7 Appendix I: Backpropagation 531
- 22.8 Appendix II: Support Vector Machines for Datasets That Are Not Linearly Separable 534
- 22.9 Appendix III: Using Support Vector Machines with Non-Linear Kernels 535

23 RECOGNITION BY RELATIONS BETWEEN TEMPLATES

537

- 23.1 Finding Objects by Voting on Relations between Templates 538
 - 23.1.1 Describing Image Patches, 538
 - 23.1.2 Voting and a Simple Generative Model, 539
 - 23.1.3 Probabilistic Models for Voting, 539
 - 23.1.4 Voting on Relations, 541
 - 23.1.5 Voting and 3D Objects, 542
- 23.2 Relational Reasoning Using Probabilistic Models and Search 542
 - 23.2.1 Correspondence and Search, 543
 - 23.2.2 Example: Finding Faces, 545
- 23.3 Using Classifiers to Prune Search 546
 - 23.3.1 Identifying Acceptable Assemblies Using Projected Classifiers, 547
 - 23.3.2 Example: Finding People and Horses Using Spatial Relations, 548
- 23.4 Technique: Hidden Markov Models 550
 - 23.4.1 Formal Matters, 550
 - 23.4.2 Computing with Hidden Markov Models, 552
 - 23.4.3 Varieties of HMMs, 557
- 23.5 Application: Hidden Markov Models and Sign Language Understanding 559
 - 23.5.1 Language Models: Sentences from Words, 560
- 23.6 Application: Finding People with Hidden Markov Models 563
- 23.7 Notes 566

24	GEOMETRIC TEMPLATES FROM SPATIAL RELATIONS	568
24.1	Simple Relations between Object and Image 568	
24.1.1	Relations for Curved Surfaces, 569	
24.1.2	Class-Based Grouping, 575	
24.2	Primitives, Templates, and Geometric Inference 576	
24.2.1	Generalized Cylinders as Volumetric Primitives, 576	
24.2.2	Ribbons, 579	
24.2.3	What Can One Represent with Ribbons? 584	
24.2.4	Linking 3D and 2D for Cylinders of Known Length, 586	
24.2.5	Linking 3D and Image Data Using Explicit Geometric Reasoning, 588	
24.3	Afterword: Object Recognition 591	
24.3.1	The Facts on the Ground, 592	
24.3.2	Current Approaches to Object Recognition, 593	
24.3.3	Limitations, 594	
24.4	Notes 594	
	Problems 595	
Part VII	Applications	597
25	APPLICATION: FINDING IN DIGITAL LIBRARIES	599
25.1	Background: Organizing Collections of Information 601	
25.1.1	How Well Does the System Work? 601	
25.1.2	What Do Users Want? 602	
25.1.3	Searching for Pictures, 602	
25.1.4	Structuring and Browsing, 604	
25.2	Summary Representations of the Whole Picture 604	
25.2.1	Histograms and Correlograms, 605	
25.2.2	Textures and Textures of Textures, 606	
25.3	Representations of Parts of the Picture 609	
25.3.1	Segmentation, 610	
25.3.2	Template Matching, 612	
25.3.3	Shape and Correspondence, 613	
25.3.4	Clustering and Organizing Collections, 614	
25.4	Video 617	
25.5	Notes 618	
26	APPLICATION: IMAGE-BASED RENDERING	620
26.1	Constructing 3D Models from Image Sequences 620	
26.1.1	Scene Modeling from Registered Images, 621	
26.1.2	Scene Modeling from Unregistered Images, 627	
26.2	Transfer-Based Approaches to Image-Based Rendering 629	
26.2.1	Affine View Synthesis, 630	
26.2.2	Euclidean View Synthesis, 632	
26.3	The Light Field 635	
26.4	Notes 639	
	Problems 639	
	BIBLIOGRAPHY	643
	INDEX	673