

Contents

I	IMAGE FORMATION	1
1	Geometric Camera Models	3
1.1	Image Formation	4
1.1.1	Pinhole Perspective	4
1.1.2	Weak Perspective	6
1.1.3	Cameras with Lenses	8
1.1.4	The Human Eye	12
1.2	Intrinsic and Extrinsic Parameters	14
1.2.1	Rigid Transformations and Homogeneous Coordinates	14
1.2.2	Intrinsic Parameters	16
1.2.3	Extrinsic Parameters	18
1.2.4	Perspective Projection Matrices	19
1.2.5	Weak-Perspective Projection Matrices	20
1.3	Geometric Camera Calibration	22
1.3.1	A Linear Approach to Camera Calibration	23
1.3.2	A Nonlinear Approach to Camera Calibration	27
1.4	Notes	29
2	Light and Shading	32
2.1	Modelling Pixel Brightness	32
2.1.1	Reflection at Surfaces	33
2.1.2	Sources and Their Effects	34
2.1.3	The Lambertian+Specular Model	36
2.1.4	Area Sources	36
2.2	Inference from Shading	37
2.2.1	Radiometric Calibration and High Dynamic Range Images	38
2.2.2	The Shape of Specularities	40
2.2.3	Inferring Lightness and Illumination	43
2.2.4	Photometric Stereo: Shape from Multiple Shaded Images	46
2.3	Modelling Interreflection	52
2.3.1	The Illumination at a Patch Due to an Area Source	52
2.3.2	Radiosity and Exitance	54
2.3.3	An Interreflection Model	55
2.3.4	Qualitative Properties of Interreflections	56
2.4	Shape from One Shaded Image	59

2.5	Notes	61
3	Color	68
3.1	Human Color Perception	68
3.1.1	Color Matching	68
3.1.2	Color Receptors	71
3.2	The Physics of Color	73
3.2.1	The Color of Light Sources	73
3.2.2	The Color of Surfaces	76
3.3	Representing Color	77
3.3.1	Linear Color Spaces	77
3.3.2	Non-linear Color Spaces	83
3.4	A Model of Image Color	86
3.4.1	The Diffuse Term	88
3.4.2	The Specular Term	90
3.5	Inference from Color	90
3.5.1	Finding Specularities Using Color	90
3.5.2	Shadow Removal Using Color	92
3.5.3	Color Constancy: Surface Color from Image Color	95
3.6	Notes	99
II	EARLY VISION: JUST ONE IMAGE	105
4	Linear Filters	107
4.1	Linear Filters and Convolution	107
4.1.1	Convolution	107
4.2	Shift Invariant Linear Systems	112
4.2.1	Discrete Convolution	113
4.2.2	Continuous Convolution	115
4.2.3	Edge Effects in Discrete Convolutions	118
4.3	Spatial Frequency and Fourier Transforms	118
4.3.1	Fourier Transforms	119
4.4	Sampling and Aliasing	121
4.4.1	Sampling	122
4.4.2	Aliasing	125
4.4.3	Smoothing and Resampling	126
4.5	Filters as Templates	131
4.5.1	Convolution as a Dot Product	131
4.5.2	Changing Basis	132
4.6	Technique: Normalized Correlation and Finding Patterns	132

4.6.1	Controlling the Television by Finding Hands by Normalized Correlation	133
4.7	Technique: Scale and Image Pyramids	134
4.7.1	The Gaussian Pyramid	135
4.7.2	Applications of Scaled Representations	136
4.8	Notes	137
5	Local Image Features	141
5.1	Computing the Image Gradient	141
5.1.1	Derivative of Gaussian Filters	142
5.2	Representing the Image Gradient	144
5.2.1	Gradient-Based Edge Detectors	145
5.2.2	Orientations	147
5.3	Finding Corners and Building Neighborhoods	148
5.3.1	Finding Corners	149
5.3.2	Using Scale and Orientation to Build a Neighborhood	151
5.4	Describing Neighborhoods with SIFT and HOG Features	155
5.4.1	SIFT Features	157
5.4.2	HOG Features	159
5.5	Computing Local Features in Practice	160
5.6	Notes	160
6	Texture	164
6.1	Local Texture Representations Using Filters	166
6.1.1	Spots and Bars	167
6.1.2	From Filter Outputs to Texture Representation	168
6.1.3	Local Texture Representations in Practice	170
6.2	Pooled Texture Representations by Discovering Textons	171
6.2.1	Vector Quantization and Textons	172
6.2.2	K-means Clustering for Vector Quantization	172
6.3	Synthesizing Textures and Filling Holes in Images	176
6.3.1	Synthesis by Sampling Local Models	176
6.3.2	Filling in Holes in Images	179
6.4	Image Denoising	182
6.4.1	Non-local Means	183
6.4.2	Block Matching 3D (BM3D)	183
6.4.3	Learned Sparse Coding	184
6.4.4	Results	186
6.5	Shape from Texture	187
6.5.1	Shape from Texture for Planes	187
6.5.2	Shape from Texture for Curved Surfaces	190

6.6	Notes	191
-----	-------	-----

III EARLY VISION: MULTIPLE IMAGES 195

7 Stereopsis 197

7.1	Binocular Camera Geometry and the Epipolar Constraint	198
7.1.1	Epipolar Geometry	198
7.1.2	The Essential Matrix	200
7.1.3	The Fundamental Matrix	201
7.2	Binocular Reconstruction	201
7.2.1	Image Rectification	202
7.3	Human Stereopsis	203
7.4	Local Methods for Binocular Fusion	205
7.4.1	Correlation	205
7.4.2	Multi-Scale Edge Matching	207
7.5	Global Methods for Binocular Fusion	210
7.5.1	Ordering Constraints and Dynamic Programming	210
7.5.2	Smoothness and Graphs	211
7.6	Using More Cameras	214
7.7	Application: Robot Navigation	215
7.8	Notes	216

8 Structure from Motion 221

8.1	Internally Calibrated Perspective Cameras	221
8.1.1	Natural Ambiguity of the Problem	223
8.1.2	Euclidean Structure and Motion from Two Images	224
8.1.3	Euclidean Structure and Motion from Multiple Images	228
8.2	Uncalibrated Weak-Perspective Cameras	230
8.2.1	Natural Ambiguity of the Problem	231
8.2.2	Affine Structure and Motion from Two Images	233
8.2.3	Affine Structure and Motion from Multiple Images	237
8.2.4	From Affine to Euclidean Shape	238
8.3	Uncalibrated Perspective Cameras	240
8.3.1	Natural Ambiguity of the Problem	241
8.3.2	Projective Structure and Motion from Two Images	242
8.3.3	Projective Structure and Motion from Multiple Images	244
8.3.4	From Projective to Euclidean Shape	246
8.4	Notes	248

IV	MID-LEVEL VISION	253
9	Segmentation by Clustering	255
9.1	Human Vision: Grouping and Gestalt	256
9.2	Important Applications	261
9.2.1	Background Subtraction	261
9.2.2	Shot Boundary Detection	264
9.2.3	Interactive Segmentation	265
9.2.4	Forming Image Regions	266
9.3	Image Segmentation by Clustering Pixels	268
9.3.1	Basic Clustering Methods	269
9.3.2	The Watershed Algorithm	271
9.3.3	Segmentation Using K-means	272
9.3.4	Mean Shift: Finding Local Modes in Data	273
9.3.5	Clustering and Segmentation with Mean Shift	275
9.4	Segmentation, Clustering, and Graphs	277
9.4.1	Terminology and Facts for Graphs	277
9.4.2	Agglomerative Clustering with a Graph	279
9.4.3	Divisive Clustering with a Graph	281
9.4.4	Normalized Cuts	284
9.5	Image Segmentation in Practice	285
9.5.1	Evaluating Segmenters	286
9.6	Notes	287
10	Grouping and Model Fitting	290
10.1	The Hough Transform	290
10.1.1	Fitting Lines with the Hough Transform	290
10.1.2	Using the Hough Transform	292
10.2	Fitting Lines and Planes	293
10.2.1	Fitting a Single Line	294
10.2.2	Fitting Planes	295
10.2.3	Fitting Multiple Lines	296
10.3	Fitting Curved Structures	297
10.4	Robustness	299
10.4.1	M-Estimators	300
10.4.2	RANSAC: Searching for Good Points	302
10.5	Fitting Using Probabilistic Models	306
10.5.1	Missing Data Problems	307
10.5.2	Mixture Models and Hidden Variables	309
10.5.3	The EM Algorithm for Mixture Models	310
10.5.4	Difficulties with the EM Algorithm	312

10.6	Motion Segmentation by Parameter Estimation	313
10.6.1	Optical Flow and Motion	315
10.6.2	Flow Models	316
10.6.3	Motion Segmentation with Layers	317
10.7	Model Selection: Which Model Is the Best Fit?	319
10.7.1	Model Selection Using Cross-Validation	322
10.8	Notes	322
11	Tracking	326
11.1	Simple Tracking Strategies	327
11.1.1	Tracking by Detection	327
11.1.2	Tracking Translations by Matching	330
11.1.3	Using Affine Transformations to Confirm a Match	332
11.2	Tracking Using Matching	334
11.2.1	Matching Summary Representations	335
11.2.2	Tracking Using Flow	337
11.3	Tracking Linear Dynamical Models with Kalman Filters	339
11.3.1	Linear Measurements and Linear Dynamics	340
11.3.2	The Kalman Filter	344
11.3.3	Forward-backward Smoothing	345
11.4	Data Association	349
11.4.1	Linking Kalman Filters with Detection Methods	349
11.4.2	Key Methods of Data Association	350
11.5	Particle Filtering	350
11.5.1	Sampled Representations of Probability Distributions	351
11.5.2	The Simplest Particle Filter	355
11.5.3	The Tracking Algorithm	356
11.5.4	A Workable Particle Filter	358
11.5.5	Practical Issues in Particle Filters	360
11.6	Notes	362
V	HIGH-LEVEL VISION	365
12	Registration	367
12.1	Registering Rigid Objects	368
12.1.1	Iterated Closest Points	368
12.1.2	Searching for Transformations via Correspondences	369
12.1.3	Application: Building Image Mosaics	370
12.2	Model-based Vision: Registering Rigid Objects with Projection	375

12.2.1	Verification: Comparing Transformed and Rendered Source to Target	377
12.3	Registering Deformable Objects	378
12.3.1	Deforming Texture with Active Appearance Models	378
12.3.2	Active Appearance Models in Practice	381
12.3.3	Application: Registration in Medical Imaging Systems	383
12.4	Notes	388
13	Smooth Surfaces and Their Outlines	391
13.1	Elements of Differential Geometry	393
13.1.1	Curves	393
13.1.2	Surfaces	397
13.2	Contour Geometry	402
13.2.1	The Occluding Contour and the Image Contour	402
13.2.2	The Cusps and Inflections of the Image Contour	403
13.2.3	Koenderink's Theorem	404
13.3	Visual Events: More Differential Geometry	407
13.3.1	The Geometry of the Gauss Map	407
13.3.2	Asymptotic Curves	409
13.3.3	The Asymptotic Spherical Map	410
13.3.4	Local Visual Events	412
13.3.5	The Bitangent Ray Manifold	413
13.3.6	Multilocal Visual Events	414
13.3.7	The Aspect Graph	416
13.4	Notes	417
14	Range Data	422
14.1	Active Range Sensors	422
14.2	Range Data Segmentation	424
14.2.1	Elements of Analytical Differential Geometry	424
14.2.2	Finding Step and Roof Edges in Range Images	426
14.2.3	Segmenting Range Images into Planar Regions	431
14.3	Range Image Registration and Model Acquisition	432
14.3.1	Quaternions	433
14.3.2	Registering Range Images	434
14.3.3	Fusing Multiple Range Images	436
14.4	Object Recognition	438
14.4.1	Matching Using Interpretation Trees	438
14.4.2	Matching Free-Form Surfaces Using Spin Images	441
14.5	Kinect	446
14.5.1	Features	447

14.5.2	Technique: Decision Trees and Random Forests	448
14.5.3	Labeling Pixels	450
14.5.4	Computing Joint Positions	453
14.6	Notes	453
15	Learning to Classify	457
15.1	Classification, Error, and Loss	457
15.1.1	Using Loss to Determine Decisions	457
15.1.2	Training Error, Test Error, and Overfitting	459
15.1.3	Regularization	460
15.1.4	Error Rate and Cross-Validation	463
15.1.5	Receiver Operating Curves	465
15.2	Major Classification Strategies	467
15.2.1	Example: Mahalanobis Distance	467
15.2.2	Example: Class-Conditional Histograms and Naive Bayes	468
15.2.3	Example: Classification Using Nearest Neighbors	469
15.2.4	Example: The Linear Support Vector Machine	470
15.2.5	Example: Kernel Machines	473
15.2.6	Example: Boosting and Adaboost	475
15.3	Practical Methods for Building Classifiers	475
15.3.1	Manipulating Training Data to Improve Performance	477
15.3.2	Building Multi-Class Classifiers Out of Binary Classifiers	479
15.3.3	Solving for SVMs and Kernel Machines	480
15.4	Notes	481
16	Classifying Images	482
16.1	Building Good Image Features	482
16.1.1	Example Applications	482
16.1.2	Encoding Layout with GIST Features	485
16.1.3	Summarizing Images with Visual Words	487
16.1.4	The Spatial Pyramid Kernel	489
16.1.5	Dimension Reduction with Principal Components	493
16.1.6	Dimension Reduction with Canonical Variates	494
16.1.7	Example Application: Identifying Explicit Images	498
16.1.8	Example Application: Classifying Materials	502
16.1.9	Example Application: Classifying Scenes	502
16.2	Classifying Images of Single Objects	504
16.2.1	Image Classification Strategies	505
16.2.2	Evaluating Image Classification Systems	505
16.2.3	Fixed Sets of Classes	508
16.2.4	Large Numbers of Classes	509

16.2.5	Flowers, Leaves, and Birds: Some Specialized Problems . . .	511
16.3	Image Classification in Practice	512
16.3.1	Codes for Image Features	513
16.3.2	Image Classification Datasets	513
16.3.3	Dataset Bias	515
16.3.4	Crowdsourcing Dataset Collection	515
16.4	Notes	517
17	Detecting Objects in Images	519
17.1	The Sliding Window Method	519
17.1.1	Face Detection	520
17.1.2	Detecting Humans	525
17.1.3	Detecting Boundaries	527
17.2	Detecting Deformable Objects	530
17.3	The State of the Art of Object Detection	535
17.3.1	Datasets and Resources	538
17.4	Notes	539
18	Topics in Object Recognition	540
18.1	What Should Object Recognition Do?	540
18.1.1	What Should an Object Recognition System Do?	540
18.1.2	Current Strategies for Object Recognition	542
18.1.3	What Is Categorization?	542
18.1.4	Selection: What Should Be Described?	544
18.2	Feature Questions	544
18.2.1	Improving Current Image Features	544
18.2.2	Other Kinds of Image Feature	546
18.3	Geometric Questions	547
18.4	Semantic Questions	549
18.4.1	Attributes and the Unfamiliar	550
18.4.2	Parts, Poselets and Consistency	551
18.4.3	Chunks of Meaning	554
VI	APPLICATIONS AND TOPICS	557
19	Image-Based Modeling and Rendering	559
19.1	Visual Hulls	559
19.1.1	Main Elements of the Visual Hull Model	561
19.1.2	Tracing Intersection Curves	563
19.1.3	Clipping Intersection Curves	566

19.1.4	Triangulating Cone Strips	567
19.1.5	Results	568
19.1.6	Going Further: Carved Visual Hulls	572
19.2	Patch-Based Multi-View Stereopsis	573
19.2.1	Main Elements of the PMVS Model	575
19.2.2	Initial Feature Matching	578
19.2.3	Expansion	579
19.2.4	Filtering	580
19.2.5	Results	581
19.3	The Light Field	584
19.4	Notes	587
20	Looking at People	590
20.1	HMM's, Dynamic Programming, and Tree-Structured Models	590
20.1.1	Hidden Markov Models	590
20.1.2	Inference for an HMM	592
20.1.3	Fitting an HMM with EM	597
20.1.4	Tree-Structured Energy Models	600
20.2	Parsing People in Images	602
20.2.1	Parsing with Pictorial Structure Models	602
20.2.2	Estimating the Appearance of Clothing	604
20.3	Tracking People	606
20.3.1	Why Human Tracking Is Hard	606
20.3.2	Kinematic Tracking by Appearance	608
20.3.3	Kinematic Human Tracking Using Templates	609
20.4	3D from 2D: Lifting	611
20.4.1	Reconstruction in an Orthographic View	611
20.4.2	Exploiting Appearance for Unambiguous Reconstructions	613
20.4.3	Exploiting Motion for Unambiguous Reconstructions	615
20.5	Activity Recognition	617
20.5.1	Background: Human Motion Data	617
20.5.2	Body Configuration and Activity Recognition	621
20.5.3	Recognizing Human Activities with Appearance Features	622
20.5.4	Recognizing Human Activities with Compositional Models	624
20.6	Resources	624
20.7	Notes	626
21	Image Search and Retrieval	627
21.1	The Application Context	627
21.1.1	Applications	628
21.1.2	User Needs	629

21.1.3	Types of Image Query	630
21.1.4	What Users Do with Image Collections	631
21.2	Basic Technologies from Information Retrieval	632
21.2.1	Word Counts	632
21.2.2	Smoothing Word Counts	633
21.2.3	Approximate Nearest Neighbors and Hashing	634
21.2.4	Ranking Documents	638
21.3	Images as Documents	639
21.3.1	Matching Without Quantization	640
21.3.2	Ranking Image Search Results	641
21.3.3	Browsing and Layout	643
21.3.4	Laying Out Images for Browsing	644
21.4	Predicting Annotations for Pictures	645
21.4.1	Annotations from Nearby Words	646
21.4.2	Annotations from the Whole Image	646
21.4.3	Predicting Correlated Words with Classifiers	648
21.4.4	Names and Faces	649
21.4.5	Generating Tags with Segments	651
21.5	The State of the Art of Word Prediction	654
21.5.1	Resources	655
21.5.2	Comparing Methods	655
21.5.3	Open Problems	656
21.6	Notes	659

VII BACKGROUND MATERIAL 661

22	Optimization Techniques	663
22.1	Linear Least-Squares Methods	663
22.1.1	Normal Equations and the Pseudoinverse	664
22.1.2	Homogeneous Systems and Eigenvalue Problems	665
22.1.3	Generalized Eigenvalues Problems	666
22.1.4	An Example: Fitting a Line to Points in a Plane	666
22.1.5	Singular Value Decomposition	667
22.2	Nonlinear Least-Squares Methods	669
22.2.1	Newton's Method: Square Systems of Nonlinear Equations.	670
22.2.2	Newton's Method for Overconstrained Systems	670
22.2.3	The Gauss–Newton and Levenberg–Marquardt Algorithms	671
22.3	Sparse Coding and Dictionary Learning	672
22.3.1	Sparse Coding	672
22.3.2	Dictionary Learning	673

22.3.3 Supervised Dictionary Learning	675
22.4 Min-Cut/Max-Flow Problems and Combinatorial Optimization . . .	675
22.4.1 Min-Cut Problems	676
22.4.2 Quadratic Pseudo-Boolean Functions	677
22.4.3 Generalization to Integer Variables	679
22.5 Notes	682
Bibliography	684
Index	737
List of Algorithms	760