
CONTENTS

PREFACE	xix
I IMAGE FORMATION	1
1 CAMERAS	3
1.1 Pinhole Cameras	4
1.1.1 Perspective Projection	4
1.1.2 Affine Projection	6
1.1.3 Spherical Projection	7
1.2 Cameras with Lenses	9
1.2.1 First-Order Geometric Optics	10
1.2.2 Thin Lenses: Geometry	12
1.2.3 Thin Lenses: Radiometry	14
1.2.4 Real Lenses	16
Human Vision: The Structure of the Eye	20
1.3 Sensing	22
1.3.1 CCD cameras	23
1.3.2 Sensor Models	24
1.4 Notes	25
1.5 Assignments	26
2 RADIOMETRY — MEASURING LIGHT	27
2.1 Light in Space	27
2.1.1 Foreshortening	27
2.1.2 Solid Angle	28
2.1.3 Radiance	30

2.2	Light at Surfaces	32
2.2.1	Simplifying Assumptions	33
2.2.2	The Bidirectional Reflectance Distribution Function	33
2.3	Important Special Cases	35
2.3.1	Radiosity	35
2.3.2	Directional Hemispheric Reflectance	36
2.3.3	Lambertian Surfaces and Albedo	36
2.3.4	Specular Surfaces	37
2.3.5	The Lambertian + Specular Model	38
2.4	Quick Reference: Radiometric Terminology for Light	40
2.5	Quick Reference: Radiometric Properties of Surfaces	41
2.6	Quick Reference: Important Types of Surface	42
2.7	Notes	43
2.8	Assignments	43
3	SOURCES, SHADOWS AND SHADING	45
3.1	Radiometric Properties of Light Sources	45
3.2	Qualitative Radiometry	46
3.3	Sources and their Effects	47
3.3.1	Point Sources	48
3.3.2	Line Sources	50
3.3.3	Area Sources	51
3.4	Local Shading Models	52
3.4.1	Local Shading Models for Point Sources	52
3.4.2	Area Sources and their Shadows	55
3.4.3	Ambient Illumination	55
3.5	Application: Photometric Stereo	57
3.5.1	Normal and Albedo from Many Views	60
3.5.2	Shape from Normals	61
3.6	Interreflections: Global Shading Models	64
3.6.1	An Interreflection Model	66
3.6.2	Solving for Radiosity	67
3.6.3	The qualitative effects of interreflections	69
3.7	Notes	71
3.8	Assignments	74
3.8.1	Exercises	74
3.8.2	Programming Assignments	75

4 COLOUR	77
4.1 Spectral Quantities	77
4.1.1 The Colour of Surfaces	77
4.1.2 The Colour of Natural Sources	79
4.1.3 Artificial Sources	79
4.2 Human Colour Perception	81
4.2.1 Colour Matching	81
4.2.2 Colour Receptors	84
4.3 Representing Colour	85
4.3.1 Linear Colour Spaces	86
4.3.2 Hue, Saturation and Value	90
4.3.3 Uniform Colour Spaces	92
4.3.4 Chromatic Adaptation	94
4.4 Application: Finding Skin Using Image Colour	94
4.4.1 The Appearance of Skin	94
4.4.2 Simple Skin Finding	95
4.5 Surface Colour from Image Colour	98
4.5.1 Surface Colour Perception in People	100
4.5.2 Inferring Lightness	102
4.5.3 A Model for Image Colour	105
4.5.4 Surface Colour from Finite Dimensional Linear Models	107
4.6 Digression: Device-Independent Colour Imaging	110
4.7 Notes	113
4.7.1 Trichromacy and Colour Spaces	113
4.7.2 Lightness and Colour Constancy	113
4.7.3 Colour in Recognition	114
4.8 Assignments	114
II IMAGE MODELS	118
5 GEOMETRIC IMAGE FEATURES	120
5.1 Elements of Differential Geometry	124
5.1.1 Curves	124
5.1.2 Surfaces	129
Application: The shape of specularities	133
5.2 Contour Geometry	136
5.2.1 The Occluding Contour and the Image Contour	137

5.2.2	The Cusps and Inflections of the Image Contour	138
5.2.3	Koenderink's Theorem	139
5.3	Notes	141
5.4	Assignments	142
6	ANALYTICAL IMAGE FEATURES	144
6.1	Elements of Analytical Euclidean Geometry	144
6.1.1	Coordinate Systems and Homogeneous Coordinates	145
6.1.2	Coordinate System Changes and Rigid Transformations	148
6.2	Geometric Camera Parameters	153
6.2.1	Intrinsic Parameters	153
6.2.2	Extrinsic Parameters	156
6.2.3	A Characterization of Perspective Projection Matrices	156
6.3	Calibration Methods	157
6.3.1	A Linear Approach to Camera Calibration	158
Technique: Linear Least Squares Methods		159
6.3.2	Taking Radial Distortion into Account	163
6.3.3	Using Straight Lines for Calibration	165
6.3.4	Analytical Photogrammetry	167
Technique: Non-Linear Least Squares Methods		169
6.4	Notes	171
6.5	Assignments	171
7	AN INTRODUCTION TO PROBABILITY	174
7.1	Probability in Discrete Spaces	175
7.1.1	Probability: the P-function	175
7.1.2	Conditional Probability	177
7.1.3	Choosing P	177
7.2	Probability in Continuous Spaces	183
7.2.1	Event Structures for Continuous Spaces	183
7.2.2	Representing a P-function for the Real Line	184
7.2.3	Probability Densities	185
7.3	Random Variables	185
7.3.1	Conditional Probability and Independence	186
7.3.2	Expectations	187
7.3.3	Joint Distributions and Marginalization	189
7.4	Standard Distributions and Densities	189
7.4.1	The Normal Distribution	191

7.5	Probabilistic Inference	191
7.5.1	The Maximum Likelihood Principle	192
7.5.2	Priors, Posteriors and Bayes' rule	194
7.5.3	Bayesian Inference	194
7.5.4	Open Issues	201
7.6	Discussion	202
III EARLY VISION: ONE IMAGE		204
8 LINEAR FILTERS		206
8.1	Linear Filters and Convolution	206
8.1.1	Convolution	206
8.1.2	Example: Smoothing by Averaging	207
8.1.3	Example: Smoothing with a Gaussian	207
8.2	Shift invariant linear systems	211
8.2.1	Discrete Convolution	212
8.2.2	Continuous Convolution	213
8.2.3	Edge Effects in Discrete Convolutions	216
8.3	Spatial Frequency and Fourier Transforms	217
8.3.1	Fourier Transforms	217
8.3.2	Differentiation and Noise	220
8.4	Sampling and Aliasing	221
8.4.1	Sampling	221
8.4.2	Aliasing	224
8.4.3	Smoothing and Resampling	226
8.5	Human: Filters and Primate Early Vision	228
8.5.1	The Visual Pathway	228
8.5.2	Experimental Methods	232
8.5.3	The Response of Retinal Cells	233
8.6	Laplacians and edges	235
8.7	Discussion	238
8.7.1	Real Imaging Systems vs Shift-Invariant Linear Systems	238
8.7.2	Scale	238
9 EDGE DETECTION		240
9.1	Estimating Derivatives with Finite Differences	240
9.2	Noise	242

9.2.1	Additive Stationary Gaussian Noise	242
9.3	Edges and Gradient-based Edge Detectors	246
9.3.1	Estimating Gradients	247
9.3.2	Choosing a Smoothing Filter	249
9.3.3	Why Smooth with a Gaussian?	251
9.3.4	Derivative of Gaussian Filters	253
9.3.5	Identifying Edge Points from Filter Outputs	254
9.4	Technique: Scale and Image Pyramids	256
9.4.1	The Gaussian Pyramid	257
9.4.2	Applications of Scaled Representations	259
9.4.3	Scale Space	260
9.4.4	Anisotropic Scaling	262
9.5	Human Vision: More of the Visual Pathway	264
9.5.1	The Lateral Geniculate Nucleus	264
9.5.2	The Visual Cortex	264
9.5.3	A Model of Early Spatial Vision	266
9.6	Commentary	268
10 NON-LINEAR FILTERS		273
10.1	Filters as Templates	273
10.1.1	Convolution as a Dot Product	274
10.1.2	Changing Basis	274
10.1.3	Normalised Correlation	275
10.1.4	Controlling the Television by Finding Hands by Normalised Correlation	275
10.1.5	More Noise Models	276
10.1.6	Spatial Noise Models	276
10.2	Non-linear Filters	278
10.2.1	Robust Estimates	278
10.2.2	Median Filters	279
10.2.3	Mathematical morphology: erosion and dilation	282
10.3	Corners and orientation representations	283
10.4	Anisotropic Scaling	285
11 TEXTURE		287
11.1	Representing Texture	289
11.1.1	Extracting Image Structure with Filter Banks	289
11.2	Analysis (and Synthesis) Using Oriented Pyramids	294

11.2.1 The Laplacian Pyramid	295
11.2.2 Oriented Pyramids	298
11.3 Application: Synthesizing Textures for Rendering	300
11.3.1 Homogeneity	300
11.3.2 Synthesis by Matching Histograms of Filter Responses	300
11.3.3 Synthesis by Sampling Conditional Densities of Filter Responses	305
11.3.4 Synthesis by Sampling Local Models	308
11.4 Shape from Texture: Planes and Isotropy	310
11.4.1 Recovering the Orientation of a Plane from an Isotropic Texture	311
11.5 Shape from Texture: Curved Surfaces	315
11.5.1 Shape from Texture for Cylinders	315
11.5.2 Obtaining the Transformation between Image Patches	324
11.6 Notes	324
11.6.1 Shape from Texture	324
IV EARLY VISION: MULTIPLE IMAGES	327
12 THE GEOMETRY OF MULTIPLE VIEWS	329
12.1 Two Views	330
12.1.1 Epipolar Geometry	330
12.1.2 The Calibrated Case	331
12.1.3 Small Motions	332
12.1.4 The Uncalibrated Case	333
12.1.5 Weak Calibration	334
12.2 Three Views	337
12.2.1 Trifocal Geometry	339
12.2.2 The Calibrated Case	339
12.2.3 The Uncalibrated Case	341
12.2.4 Estimation of the Trifocal Tensor	342
12.3 More Views	343
12.4 Notes	349
12.5 Assignments	351
13 STEREOPSIS	353
13.1 Reconstruction	355
13.1.1 Camera Calibration	356
13.1.2 Image Rectification	357

Human Vision: Stereopsis	359
13.2 Binocular Fusion	363
13.2.1 Correlation	363
13.2.2 Multi-Scale Edge Matching	365
13.2.3 Dynamic Programming	368
13.3 Using More Cameras	370
13.3.1 Trinocular Stereo	370
13.3.2 Multiple-Baseline Stereo	372
13.4 Notes	373
13.5 Assignments	375
14 AFFINE STRUCTURE FROM MOTION	378
14.1 Elements of Affine Geometry	379
14.2 Affine Structure from Two Images	382
14.2.1 The Affine Structure-from-Motion Theorem	383
14.2.2 Rigidity and Metric Constraints	383
14.3 Affine Structure from Multiple Images	385
14.3.1 The Affine Structure of Affine Image Sequences	386
Technique: Singular Value Decomposition	386
14.3.2 A Factorization Approach to Affine Motion Analysis	388
14.4 From Affine to Euclidean Images	389
14.4.1 Euclidean Projection Models	390
14.4.2 From Affine to Euclidean Motion	391
14.5 Affine Motion Segmentation	392
14.5.1 The Reduced Echelon Form of the Data Matrix	392
14.5.2 The Shape Interaction Matrix	393
14.6 Notes	395
14.7 Assignments	396
15 PROJECTIVE STRUCTURE FROM MOTION	398
15.1 Elements of Projective Geometry	399
15.1.1 Projective Bases and Projective Coordinates	399
15.1.2 Projective Transformations	401
15.1.3 Affine and Projective Spaces	403
15.1.4 Hyperplanes and Duality	404
15.1.5 Cross-Ratios	405
15.1.6 Application: Parameterizing the Fundamental Matrix	408
15.2 Projective Scene Reconstruction from Two Views	409

15.2.1 Analytical Scene Reconstruction	409
15.2.2 Geometric Scene Reconstruction	411
15.3 Motion Estimation from Two or Three Views	412
15.3.1 Motion Estimation from Fundamental Matrices	412
15.3.2 Motion Estimation from Trifocal Tensors	414
15.4 Motion Estimation from Multiple Views	415
15.4.1 A Factorization Approach to Projective Motion Analysis	416
15.4.2 Bundle Adjustment	419
15.5 From Projective to Euclidean Images	419
15.5.1 Computing Euclidean Upgrades under Minimal Assumptions	421
15.6 Notes	424
15.7 Assignments	426
V MID-LEVEL VISION	431
16 SEGMENTATION USING CLUSTERING METHODS	433
16.1 Human vision: Grouping and Gestalt	435
16.2 Application: Shot Boundary Detection and Background Subtraction	439
16.2.1 Background Subtraction	439
16.2.2 Shot Boundary Detection	441
16.3 Image Segmentation by Clustering Pixels	442
16.3.1 Simple Clustering Methods	442
16.3.2 Segmentation Using Simple Clustering Methods	444
16.3.3 Clustering and Segmentation by K-means	447
16.4 Segmentation by Graph-Theoretic Clustering	448
16.4.1 Basic Graphs	449
16.4.2 The Overall Approach	450
16.4.3 Affinity Measures	452
16.4.4 Eigenvectors and Segmentation	455
16.4.5 Normalised Cuts	458
16.5 Discussion	462
17 FITTING	467
17.1 The Hough Transform	468
17.1.1 Fitting Lines with the Hough Transform	468
17.1.2 Practical Problems with the Hough Transform	469
17.2 Fitting Lines	471

17.2.1 Least Squares, Maximum Likelihood and Parameter Estimation	472
17.2.2 Which Point is on Which Line?	475
17.3 Fitting Curves	476
17.3.1 Implicit Curves	477
17.3.2 Parametric Curves	480
17.4 Fitting to the Outlines of Surfaces	481
17.4.1 Some Relations Between Surfaces and Outlines	482
17.4.2 Clustering to Form Symmetries	484
17.5 Discussion	488
18 SEGMENTATION AND FITTING USING PROBABILISTIC METHODS	491
18.1 Missing Data Problems, Fitting and Segmentation	492
18.1.1 Missing Data Problems	492
18.1.2 The EM Algorithm	494
18.1.3 Colour and Texture Segmentation with EM	500
18.1.4 Motion Segmentation and EM	501
18.1.5 The Number of Components	505
18.1.6 How Many Lines are There?	505
18.2 Robustness	506
18.2.1 Explicit Outliers	506
18.2.2 M-estimators	508
18.2.3 RANSAC	511
18.3 How Many are There?	514
18.3.1 Basic Ideas	515
18.3.2 AIC — An Information Criterion	515
18.3.3 Bayesian methods and Schwartz' BIC	516
18.3.4 Description Length	517
18.3.5 Other Methods for Estimating Deviance	517
18.4 Discussion	518
19 TRACKING	520
19.1 Tracking as an Abstract Inference Problem	521
19.1.1 Independence Assumptions	521
19.1.2 Tracking as Inference	522
19.1.3 Overview	523
19.2 Linear Dynamic Models and the Kalman Filter	523
19.2.1 Linear Dynamic Models	523

19.2.2	Kalman Filtering	528
19.2.3	The Kalman Filter for a 1D State Vector	528
19.2.4	The Kalman Update Equations for a General State Vector	530
19.2.5	Forward-Backward Smoothing	531
19.3	Non-Linear Dynamic Models	536
19.3.1	Unpleasant Properties of Non-Linear Dynamics	539
19.3.2	Difficulties with Likelihoods	540
19.4	Particle Filtering	542
19.4.1	Sampled Representations of Probability Distributions	542
19.4.2	The Simplest Particle Filter	546
19.4.3	A Workable Particle Filter	549
19.4.4	If's, And's and But's — Practical Issues in Building Particle Filters	550
19.5	Data Association	554
19.5.1	Choosing the Nearest — Global Nearest Neighbours	554
19.5.2	Gating and Probabilistic Data Association	555
19.6	Applications and Examples	558
19.6.1	Vehicle Tracking	559
19.6.2	Finding and Tracking People	563
19.7	Discussion	569
II	Appendix: The Extended Kalman Filter, or EKF	572
VI	HIGH-LEVEL VISION	574
20	CORRESPONDENCE AND POSE CONSISTENCY	576
20.1	Initial Assumptions	576
20.1.1	Obtaining Hypotheses	577
20.2	Obtaining Hypotheses by Pose Consistency	578
20.2.1	Pose Consistency for Perspective Cameras	579
20.2.2	Affine and Projective Camera Models	581
20.2.3	Linear Combinations of Models	583
20.3	Obtaining Hypotheses by Pose Clustering	585
20.4	Obtaining Hypotheses Using Invariants	586
20.4.1	Invariants for Plane Figures	588
20.4.2	Geometric Hashing	591
20.4.3	Invariants and Indexing	592
20.5	Verification	596

20.5.1	Edge Proximity	597
20.5.2	Similarity in Texture, Pattern and Intensity	599
20.5.3	Example: Bayes Factors and Verification	599
20.6	Application: Registration in Medical Imaging Systems	600
20.6.1	Imaging Modes	601
20.6.2	Applications of Registration	602
20.6.3	Geometric Hashing Techniques in Medical Imaging	603
20.7	Curved Surfaces and Alignment	605
20.8	Discussion	608
21	FINDING TEMPLATES USING CLASSIFIERS	613
21.1	Classifiers	614
21.1.1	Using Loss to Determine Decisions	614
21.1.2	Overview: Methods for Building Classifiers	616
21.1.3	Example: A Plug-in Classifier for Normal Class-conditional Densities	618
21.1.4	Example: A Non-Parametric Classifier using Nearest Neighbours	619
21.1.5	Estimating and Improving Performance	620
21.2	Building Classifiers from Class Histograms	622
21.2.1	Finding Skin Pixels using a Classifier	623
21.2.2	Face Finding Assuming Independent Template Responses	624
21.3	Feature Selection	627
21.3.1	Principal Component Analysis	627
21.3.2	Canonical Variates	629
21.4	Neural Networks	633
21.4.1	Key Ideas	633
21.4.2	Minimizing the Error	638
21.4.3	When to Stop Training	642
21.4.4	Finding Faces using Neural Networks	642
21.4.5	Convolutional Neural Nets	644
21.5	The Support Vector Machine	647
21.5.1	Support Vector Machines for Linearly Separable Datasets	648
21.5.2	Finding Pedestrians using Support Vector Machines	650
21.6	Conclusions	654
II	Appendix: Support Vector Machines for Datasets that are not Linearly Separable	656
III	Appendix: Using Support Vector Machines with Non-Linear Kernels	657

22 RECOGNITION BY RELATIONS BETWEEN TEMPLATES	659
22.1 Relational Reasoning using Probabilistic Models and Search	660
22.1.1 Growing Assemblies Incrementally	660
22.1.2 Detection	662
22.1.3 Pruning	663
22.2 Frames and Probability Models	664
22.2.1 Representing Coordinate Frames Explicitly in a Probability Model	664
22.2.2 Using a Probability Model to Predict Feature Positions	665
22.2.3 Building Probability Models that are Frame-Invariant	668
22.2.4 Example: Finding Faces Using Frame Invariance	668
22.3 Finding Objects by Voting on Relations between Templates	669
22.3.1 Interest Points	669
22.3.2 Simple voting	671
22.3.3 Voting on Relations	671
22.3.4 Voting and 3D Objects	673
22.4 Using Classifiers to Prune Search	674
22.4.1 Identifying Acceptable Assemblies Using Projected Classifiers	674
22.4.2 Finding People and Horses Using Spatial Relations	674
22.5 Technique: Hidden Markov Models	676
22.5.1 Formal Matters	676
22.5.2 Computing with Hidden Markov Models	678
22.5.3 Varieties of HMM's	682
22.6 Sign Language, Gestures and HMM's	682
22.7 Conclusions	685
23 ASPECT GRAPHS	694
23.1 Differential Geometry and Visual Events	699
23.1.1 The Geometry of the Gauss Map	699
23.1.2 Asymptotic Curves	701
23.1.3 The Asymptotic Spherical Map	703
23.1.4 Local Visual Events	704
23.1.5 The Bitangent Ray Manifold	706
23.1.6 Multilocal Visual Events	708
23.1.7 Remarks	709
23.2 Computing the Aspect Graph	711
23.2.1 Step 1: Tracing Visual Events	712
23.2.2 Step 2: Constructing the Regions	713

23.2.3 Remaining Steps of the Algorithm	714
23.2.4 An Example	714
23.3 Aspect Graphs and Object Recognition	718
23.4 Notes	718
23.5 Assignments	720
VII APPLICATIONS AND TOPICS	721
24 RANGE DATA	723
24.1 Active Range Sensors	723
24.2 Range Data Segmentation	726
Technique: Analytical Differential Geometry	727
24.2.1 Finding Step and Roof Edges in Range Images	729
24.2.2 Segmenting Range Images into Planar Regions	734
24.3 Range Image Registration and Model Construction	736
Technique: Quaternions	737
24.3.1 Registering Range Images Using the Iterative Closest-Point Method	738
24.3.2 Fusing Multiple Range Images	741
24.4 Object Recognition	742
24.4.1 Matching Piecewise-Planar Surfaces Using Interpretation Trees	743
24.4.2 Matching Free-Form Surfaces Using Spin Images	746
24.5 Notes	751
24.6 Assignments	752
25 APPLICATION: FINDING IN DIGITAL LIBRARIES	754
25.1 Background	755
25.1.1 What do users want?	755
25.1.2 What can tools do?	757
25.2 Appearance	758
25.2.1 Histograms and correlograms	759
25.2.2 Textures and textures of textures	760
25.3 Finding	768
25.3.1 Annotation and segmentation	770
25.3.2 Template matching	771
25.3.3 Shape and correspondence	773
25.4 Video	775

25.5 Discussion	779
26 APPLICATION: IMAGE-BASED RENDERING	780
26.1 Constructing 3D Models from Image Sequences	781
26.1.1 Scene Modeling from Registered Images	781
26.1.2 Scene Modeling from Unregistered Images	790
26.2 Transfer-Based Approaches to Image-Based Rendering	794
26.2.1 Affine View Synthesis	796
26.2.2 Euclidean View Synthesis	798
26.3 The Light Field	801
26.4 Notes	805
26.5 Assignments	806
BIBLIOGRAPHY	809