Contents

I IMAGE FORMATION

| Geo | ometric | c Camera Models | 33 |
|-----|---|--|--|
| 1.1 | Image | Formation | 34 |
| | 1.1.1 | Pinhole Perspective | 34 |
| | 1.1.2 | Weak Perspective | 36 |
| | 1.1.3 | Cameras with Lenses | 38 |
| | 1.1.4 | The Human Eye | 42 |
| 1.2 | Intrins | sic and Extrinsic Parameters | 44 |
| | 1.2.1 | Rigid Transformations and Homogeneous Coordinates $\ . \ . \ .$ | 44 |
| | 1.2.2 | Intrinsic Parameters | 46 |
| | 1.2.3 | Extrinsic Parameters | 48 |
| | 1.2.4 | Perspective Projection Matrices | 49 |
| | 1.2.5 | Weak-Perspective Projection Matrices | 50 |
| 1.3 | Geome | etric Camera Calibration | 52 |
| | 1.3.1 | A Linear Approach to Camera Calibration | 53 |
| | 1.3.2 | A Nonlinear Approach to Camera Calibration | 57 |
| 1.4 | Notes | | 59 |
| Lig | ht and | Shading | 62 |
| 2.1 | Model | ling Pixel Brightness | 62 |
| 2.1 | 2.1.1 | Reflection at Surfaces | 63 |
| | 2.1.1 | Sources and Their Effects | 64 |
| | 213 | The Lambertian+Specular Model | 66 |
| | 2.1.0 | Area Sources | 66 |
| 22 | Infere | nce from Shading | 67 |
| 2.2 | 2.2.1 | Radiometric Calibration and High Dynamic Range Images | 68 |
| | 2.2.2 | The Shape of Specularities | 70 |
| | 2.2.3 | Inferring Lightness and Illumination | 73 |
| | 2.2.4 | Photometric Stereo: Shape from Multiple Shaded Images | 76 |
| 2.3 | Model | ling Interreflection | 82 |
| - | 2.3.1 | The Illumination at a Patch Due to an Area Source | 82 |
| | | יייות אינו אינו אינו אינו אינו אינו אינו אינו | 0.4 |
| | 2.3.2 | Radiosity and Exitance | 84 |
| | $2.3.2 \\ 2.3.3$ | An Interreflection Model | $\frac{84}{85}$ |
| | $2.3.2 \\ 2.3.3 \\ 2.3.4$ | Radiosity and Exitance | 84 85 86 |
| | Ged 1.1 1.2 1.3 1.4 Lig 2.1 2.2 2.3 | $\begin{array}{c cccc} \textbf{Geometric} \\ \textbf{I.1} & Image \\ 1.1. & 1.1.2 \\ 1.1.3 \\ 1.1.2 & 1.1.3 \\ 1.1.4 \\ \textbf{I.2} & Intrins \\ 1.2.1 & 1.2.2 \\ 1.2.3 & 1.2.4 \\ 1.2.5 \\ \textbf{I.3} & Geome \\ 1.3.1 & 1.3.2 \\ \textbf{I.4} & Notes \\ \textbf{Light} & \textbf{and} \\ \textbf{2.1} & Notes \\ \textbf{Light} & \textbf{and} \\ \textbf{2.1.1} & 2.1.2 \\ 2.1.3 & 2.1.4 \\ \textbf{2.2} & Inferen \\ 2.2.1 & 2.2.2 \\ 2.2.3 & 2.2.4 \\ \textbf{2.3} & Model \\ \textbf{2.3.1} \end{array}$ | Geometric Camera Models 1.1 Image Formation 1.1.1 Pinhole Perspective 1.1.2 Weak Perspective 1.1.3 Cameras with Lenses 1.1.4 The Human Eye 1.1.4 The Human Eye 1.2 Intrinsic and Extrinsic Parameters 1.2.1 Rigid Transformations and Homogeneous Coordinates 1.2.2 Intrinsic Parameters 1.2.3 Extrinsic Parameters 1.2.4 Perspective Projection Matrices 1.2.5 Weak-Perspective Projection Matrices 1.2.6 Geometric Camera Calibration 1.3.1 A Linear Approach to Camera Calibration 1.3.2 A Nonlinear Approach to Camera Calibration 1.3.4 Notes 2.11 Reflection at Surfaces 2.12 Sources and Their Effects 2.13 The Lambertian+Specular Model 2.14 Area Sources 2.25 Inference from Shading 2.26 The Shape of Specularities 2.27 The Shape of Specularities 2.28 Inferring Lightness and Illumination 2.24 Photometric Stereo: Shape from Multiple Shaded Images 2.23 Modelling Interreflection |

 $\mathbf{31}$

| | 2.5 | Notes | |
|---|------|---------|---|
| 3 | Colo | or | 98 |
| : | 3.1 | Humar | Color Perception |
| | | 3.1.1 | Color Matching |
| | | 3.1.2 | Color Receptors |
| ; | 3.2 | The Pl | nysics of Color |
| | | 3.2.1 | The Color of Light Sources |
| | | 3.2.2 | The Color of Surfaces 106 |
| ; | 3.3 | Repres | enting Color |
| | | 3.3.1 | Linear Color Spaces |
| | | 3.3.2 | Non-linear Color Spaces |
| : | 3.4 | A Mod | lel of Image Color |
| | | 3.4.1 | The Diffuse Term |
| | | 3.4.2 | The Specular Term |
| : | 3.5 | Inferen | ce from Color |
| | | 3.5.1 | Finding Specularities Using Color |
| | | 3.5.2 | Shadow Removal Using Color |
| | | 3.5.3 | Color Constancy: Surface Color from Image Color 125 |
| | 3.6 | Notes | |

| II | EARLY | VISION: | JUST | ONE IMAGE | |
|----|-------|---------|------|-----------|--|
|----|-------|---------|------|-----------|--|

| 4 | Line | ear Filters | 137 |
|---|------|--|-----|
| | 4.1 | Linear Filters and Convolution | 137 |
| | | 4.1.1 Convolution | 137 |
| | 4.2 | Shift Invariant Linear Systems | 142 |
| | | 4.2.1 Discrete Convolution | 143 |
| | | 4.2.2 Continuous Convolution | 145 |
| | | 4.2.3 Edge Effects in Discrete Convolutions | 148 |
| | 4.3 | Spatial Frequency and Fourier Transforms | 148 |
| | | 4.3.1 Fourier Transforms | 149 |
| | 4.4 | Sampling and Aliasing | 151 |
| | | 4.4.1 Sampling | 152 |
| | | 4.4.2 Aliasing | 155 |
| | | 4.4.3 Smoothing and Resampling | 156 |
| | 4.5 | Filters as Templates | 161 |
| | | 4.5.1 Convolution as a Dot Product | 161 |
| | | 4.5.2 Changing Basis | 162 |
| | 4.6 | Technique: Normalized Correlation and Finding Patterns $\ . \ . \ .$ | 162 |
| | | | |

| | | 4.6.1 | Controlling the Television by Finding Hands by Normalized Correlation 163 |
|----------|--|---|--|
| | 4.7 | Techni | que: Scale and Image Pyramids |
| | 1.1 | 4.7.1 | The Gaussian Pyramid |
| | | 4.7.2 | Applications of Scaled Representations |
| | 4.8 | Notes | |
| | | | |
| 5 | Loc | al Ima | ge Features 171 |
| | 5.1 | Comp | uting the Image Gradient |
| | | 5.1.1 | Derivative of Gaussian Filters |
| | 5.2 | Repres | senting the Image Gradient |
| | | 5.2.1 | Gradient-Based Edge Detectors |
| | | 5.2.2 | Orientations |
| | 5.3 | Findin | g Corners and Building Neighborhoods |
| | | 5.3.1 | Finding Corners |
| | | 5.3.2 | Using Scale and Orientation to Build a Neighborhood 181 |
| | 5.4 | Descri | bing Neighborhoods with SIFT and HOG Features 185 |
| | | 5.4.1 | SIFT Features |
| | | 5.4.2 | HOG Features |
| | 5.5 | Comp | uting Local Features in Practice |
| | 5.6 | Notes | |
| | | | |
| 6 | Tex | ture | 194 |
| 6 | Tex 6.1 | ture Local | 194 Texture Representations Using Filters |
| 6 | Tex 6.1 | ture Local 6.1.1 | 194 Texture Representations Using Filters 196 Spots and Bars 197 |
| 6 | Tex 6.1 | ture Local 6.1.1 6.1.2 | 194 Texture Representations Using Filters 196 Spots and Bars 197 From Filter Outputs to Texture Representation 198 |
| 6 | Tex 6.1 | ture Local 6.1.1 6.1.2 6.1.3 | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200 |
| 6 | Tex 6.1 | ture Local 6.1.1 6.1.2 6.1.3 Pooled | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200I Texture Representations by Discovering Textons201 |
| 6 | Tex 6.1 6.2 | ture Local (6.1.1 6.1.2 6.1.3 Pooled 6.2.1 | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200I Texture Representations by Discovering Textons201Vector Quantization and Textons202 |
| 6 | Tex 6.1 6.2 | ture Local 6 6.1.1 6.1.2 6.1.3 Pooled 6.2.1 6.2.2 | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200Texture Representations by Discovering Textons201Vector Quantization and Textons202K-means Clustering for Vector Quantization202 |
| 6 | Tex 6.1 6.2 6.3 | ture Local 5 6.1.1 6.1.2 6.1.3 Pooled 6.2.1 6.2.2 Synthe | 194Texture Representations Using Filters |
| 6 | Tex 6.1 6.2 6.3 | ture Local 6 6.1.1 6.1.2 6.1.3 Pooled 6.2.1 6.2.2 Synthe 6.3.1 | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200I Texture Representations by Discovering Textons201Vector Quantization and Textons202K-means Clustering for Vector Quantization202esizing Textures and Filling Holes in Images206Synthesis by Sampling Local Models206 |
| 6 | Tex 6.1 6.2 6.3 | ture Local 6 6.1.1 6.1.2 6.1.3 Pooled 6.2.1 6.2.2 Synthe 6.3.1 6.3.2 | 194Texture Representations Using Filters |
| 6 | Tex 6.1 6.2 6.3 6.4 | ture Local 6.1.1 6.1.2 6.1.3 Pooled 6.2.1 6.2.2 Synthe 6.3.1 6.3.2 Image | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200Texture Representations by Discovering Textons201Vector Quantization and Textons202K-means Clustering for Vector Quantization202szizing Textures and Filling Holes in Images206Synthesis by Sampling Local Models209Denoising212 |
| 6 | Tex 6.1 6.2 6.3 6.4 | ture Local 4 6.1.1 6.1.2 6.1.3 Poolec 6.2.1 6.2.2 Synthe 6.3.1 6.3.2 Image 6.4.1 | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200Texture Representations by Discovering Textons201Vector Quantization and Textons202K-means Clustering for Vector Quantization202esizing Textures and Filling Holes in Images206Synthesis by Sampling Local Models209Denoising212Non-local Means213 |
| 6 | Tex 6.1 6.2 6.3 6.4 | ture Local 6.1.1 6.1.2 6.1.3 Pooled 6.2.1 6.2.2 Synthe 6.3.1 6.3.2 Image 6.4.1 6.4.2 | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200Texture Representations by Discovering Textons201Vector Quantization and Textons202K-means Clustering for Vector Quantization202esizing Textures and Filling Holes in Images206Synthesis by Sampling Local Models209Denoising212Non-local Means213Block Matching 3D (BM3D)213 |
| 6 | Tex 6.1 6.2 6.3 6.4 | ture Local 6 6.1.1 6.1.2 6.1.3 Pooled 6.2.1 6.2.2 Synthe 6.3.1 6.3.2 Image 6.4.1 6.4.2 6.4.3 | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200Texture Representations by Discovering Textons201Vector Quantization and Textons202K-means Clustering for Vector Quantization202esizing Textures and Filling Holes in Images206Synthesis by Sampling Local Models209Denoising212Non-local Means213Block Matching 3D (BM3D)213Learned Sparse Coding214 |
| 6 | Tex 6.1 6.2 6.3 6.4 | ture Local 6.1.1 6.1.2 6.1.3 Poolec 6.2.1 6.2.2 Synthe 6.3.1 6.3.2 Image 6.4.1 6.4.2 6.4.3 6.4.4 | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200Texture Representations by Discovering Textons201Vector Quantization and Textons202K-means Clustering for Vector Quantization202esizing Textures and Filling Holes in Images206Synthesis by Sampling Local Models209Denoising212Non-local Means213Block Matching 3D (BM3D)213Learned Sparse Coding214Results216 |
| 6 | Tex 6.1 6.2 6.3 6.4 | ture Local 6.1.1 6.1.2 6.1.3 Pooled 6.2.1 6.2.2 Synthe 6.3.1 6.3.2 Image 6.4.1 6.4.2 6.4.3 6.4.4 Shape | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200Texture Representations by Discovering Textons201Vector Quantization and Textons202K-means Clustering for Vector Quantization202esizing Textures and Filling Holes in Images206Synthesis by Sampling Local Models206Filling in Holes in Images202Non-local Means213Block Matching 3D (BM3D)213Learned Sparse Coding214Results216from Texture217 |
| 6 | Tex 6.1 6.2 6.3 6.4 | ture Local 6 6.1.1 6.1.2 6.1.3 Poolec 6.2.1 6.2.2 Synthe 6.3.1 6.3.2 Image 6.4.1 6.4.2 6.4.3 6.4.4 Shape 6.5.1 | 194Texture Representations Using Filters196Spots and Bars197From Filter Outputs to Texture Representation198Local Texture Representations in Practice200Texture Representations by Discovering Textons201Vector Quantization and Textons202K-means Clustering for Vector Quantization202szizing Textures and Filling Holes in Images206Synthesis by Sampling Local Models209Denoising212Non-local Means213Block Matching 3D (BM3D)213Learned Sparse Coding214Results216from Texture217Shape from Texture for Planes217 |

| 7 Stereopsis 7.1 Binocular Camera Geometry and the Epipolar Constraint | 11 1 | EARLY VISION: MULTIPLE IMAGES |
|---|------|---|
| 7.1 Binocular Camera Geometry and the Epipolar Constraint 7.1.1 Epipolar Geometry 7.1.2 The Essential Matrix 7.1.3 The Fundamental Matrix 7.1.3 The Fundamental Matrix 7.2 Binocular Reconstruction 7.2.1 Image Rectification 7.3 Human Stereopsis 7.4 Local Methods for Binocular Fusion 7.4.1 Correlation 7.4.2 Multi-Scale Edge Matching 7.5 Global Methods for Binocular Fusion 7.5.1 Ordering Constraints and Dynamic Programming 7.5.2 Smoothness and Graphs 7.6 Using More Cameras 7.7 Application: Robot Navigation 7.8 Notes 8 Structure from Motion 8.1 Internally Calibrated Perspective Cameras 8.1.1 Natural Ambiguity of the Problem 8.1.2 Euclidean Structure and Motion from Multiple Images 8.2.1 Natural Ambiguity of the Problem 8.2.2 Affine Structure and Motion from Two Images 8.2.3 Affine Structure and Motion from Multiple Images 8.3.4 | Ste | reopsis |
| 7.1.1 Epipolar Geometry 7.1.2 The Essential Matrix 7.1.3 The Fundamental Matrix 7.1.3 The Fundamental Matrix 7.2 Binocular Reconstruction 7.2 Binocular Reconstruction 7.2 Image Rectification 7.3 Human Stereopsis 7.4 Local Methods for Binocular Fusion 7.4.1 Correlation 7.4.2 Multi-Scale Edge Matching 7.5 Global Methods for Binocular Fusion 7.5.1 Ordering Constraints and Dynamic Programming 7.5.2 Smoothness and Graphs 7.6 Using More Cameras 7.7 Application: Robot Navigation 7.8 Notes 8 Structure from Motion 8.1 Internally Calibrated Perspective Cameras 8.1.1 Natural Ambiguity of the Problem 8.1.2 Euclidean Structure and Motion from Multiple Images 8.2 Uncalibrated Weak-Perspective Cameras 8.2.1 Natural Ambiguity of the Problem 8.2.2 Affine Structure and Motion from Two Images 8.2.3 Affine Structure and Motion from | 7.1 | Binocular Camera Geometry and the Epipolar Constraint $\ . \ . \ .$ |
| 7.1.2 The Essential Matrix 7.1.3 The Fundamental Matrix 7.1.3 The Fundamental Matrix 7.2 Binocular Reconstruction 7.2.1 Image Rectification 7.3 Human Stereopsis 7.4 Local Methods for Binocular Fusion 7.4.1 Correlation 7.4.2 Multi-Scale Edge Matching 7.4.2 Multi-Scale Edge Matching 7.5 Global Methods for Binocular Fusion 7.5.1 Ordering Constraints and Dynamic Programming 7.5.2 Smoothness and Graphs 7.6 Using More Cameras 7.7 Application: Robot Navigation 7.8 Notes 8 Structure from Motion 8.1 Internally Calibrated Perspective Cameras 8.1.1 Natural Ambiguity of the Problem 8.1.2 Euclidean Structure and Motion from Multiple Images 8.2.1 Natural Ambiguity of the Problem 8.2.2 Affine Structure and Motion from Two Images 8.2.3 Affine to Euclidean Shape 8.3 Uncalibrated Perspective Cameras 8.3.4 From Affine to Euclidean | | 7.1.1 Epipolar Geometry |
| 7.1.3 The Fundamental Matrix 7.2 Binocular Reconstruction 7.3 Human Stereopsis 7.4 Local Methods for Binocular Fusion 7.4.1 Correlation 7.4.2 Multi-Scale Edge Matching 7.5 Global Methods for Binocular Fusion 7.5.1 Ordering Constraints and Dynamic Programming 7.5.2 Smoothness and Graphs 7.6 Using More Cameras 7.7 Application: Robot Navigation 7.8 Notes 8 Structure from Motion 8.1 Internally Calibrated Perspective Cameras 8.1.1 Natural Ambiguity of the Problem 8.1.2 Euclidean Structure and Motion from Multiple Images 8.2.1 Natural Ambiguity of the Problem 8.2.2 Affine Structure and Motion from Two Images 8.2.3 Affine Structure and Motion from Multiple Images 8.3.4 From Affine to Euclidean Shape 8.3.1 Natural Ambiguity of the Problem | | 7.1.2 The Essential Matrix |
| 7.2 Binocular Reconstruction | | 7.1.3 The Fundamental Matrix |
| 7.2.1 Image Rectification | 7.2 | Binocular Reconstruction |
| 7.3 Human Stereopsis | | 7.2.1 Image Rectification |
| 7.4 Local Methods for Binocular Fusion | 7.3 | Human Stereopsis |
| 7.4.1 Correlation 7.4.2 Multi-Scale Edge Matching 7.5 Global Methods for Binocular Fusion 7.5.1 Ordering Constraints and Dynamic Programming 7.5.2 Smoothness and Graphs 7.6 Using More Cameras 7.7 Application: Robot Navigation 7.8 Notes 7.8 Notes 8 Structure from Motion 8.1 Internally Calibrated Perspective Cameras 8.1.1 Natural Ambiguity of the Problem 8.1.2 Euclidean Structure and Motion from Two Images 8.1.3 Euclidean Structure and Motion from Multiple Images 8.2.1 Natural Ambiguity of the Problem 8.2.2 Affine Structure and Motion from Two Images 8.2.3 Affine Structure and Motion from Multiple Images 8.2.4 From Affine to Euclidean Shape 8.3.1 Natural Ambiguity of the Problem 8.3.2 Projective Structure and Motion from Two Images | 7.4 | Local Methods for Binocular Fusion |
| 7.4.2 Multi-Scale Edge Matching 7.5 Global Methods for Binocular Fusion 7.5.1 Ordering Constraints and Dynamic Programming 7.5.2 Smoothness and Graphs 7.6 Using More Cameras 7.7 Application: Robot Navigation 7.8 Notes 7.8 Notes 8 Structure from Motion 8.1 Internally Calibrated Perspective Cameras 8.1.1 Natural Ambiguity of the Problem 8.1.2 Euclidean Structure and Motion from Two Images 8.1.3 Euclidean Structure and Motion from Multiple Images 8.2.1 Natural Ambiguity of the Problem 8.2.2 Affine Structure and Motion from Two Images 8.2.3 Affine Structure and Motion from Multiple Images 8.2.4 From Affine to Euclidean Shape 8.3 Uncalibrated Perspective Cameras 8.3.1 Natural Ambiguity of the Problem 8.3.2 Projective Structure and Motion from Two Images | | 7.4.1 Correlation |
| 7.5 Global Methods for Binocular Fusion | | 7.4.2 Multi-Scale Edge Matching |
| 7.5.1 Ordering Constraints and Dynamic Programming 7.5.2 Smoothness and Graphs | 7.5 | Global Methods for Binocular Fusion |
| 7.5.2 Smoothness and Graphs | | 7.5.1 Ordering Constraints and Dynamic Programming |
| 7.6 Using More Cameras | | 7.5.2 Smoothness and Graphs |
| 7.7 Application: Robot Navigation | 7.6 | Using More Cameras |
| 7.8 Notes | 7.7 | Application: Robot Navigation |
| 8 Structure from Motion 8.1 Internally Calibrated Perspective Cameras | 7.8 | Notes |
| 8.1 Internally Calibrated Perspective Cameras | Str | acture from Motion |
| 8.1.1 Natural Ambiguity of the Problem | 8.1 | Internally Calibrated Perspective Cameras |
| 8.1.2 Euclidean Structure and Motion from Two Images | | 8.1.1 Natural Ambiguity of the Problem |
| 8.1.3 Euclidean Structure and Motion from Multiple Images 8.2 Uncalibrated Weak-Perspective Cameras 8.2.1 Natural Ambiguity of the Problem 8.2.2 Affine Structure and Motion from Two Images 8.2.3 Affine Structure and Motion from Multiple Images 8.2.4 From Affine to Euclidean Shape 8.3 Uncalibrated Perspective Cameras 8.3.1 Natural Ambiguity of the Problem 8.3.2 Projective Structure and Motion from Two Images | | 8.1.2 Euclidean Structure and Motion from Two Images |
| 8.2 Uncalibrated Weak-Perspective Cameras | | 8.1.3 Euclidean Structure and Motion from Multiple Images |
| 8.2.1 Natural Ambiguity of the Problem | 8.2 | Uncalibrated Weak-Perspective Cameras |
| 8.2.2 Affine Structure and Motion from Two Images 8.2.3 Affine Structure and Motion from Multiple Images 8.2.4 From Affine to Euclidean Shape | | 8.2.1 Natural Ambiguity of the Problem |
| 8.2.3 Affine Structure and Motion from Multiple Images 8.2.4 From Affine to Euclidean Shape | | 8.2.2 Affine Structure and Motion from Two Images |
| 8.2.4 From Affine to Euclidean Shape | | 8.2.3 Affine Structure and Motion from Multiple Images |
| 8.3 Uncalibrated Perspective Cameras | | 8.2.4 From Affine to Euclidean Shape |
| 8.3.1 Natural Ambiguity of the Problem | 8.3 | Uncalibrated Perspective Cameras |
| 8.3.2 $$ Projective Structure and Motion from Two Images | | 8.3.1 Natural Ambiguity of the Problem |
| | | 8.3.2 Projective Structure and Motion from Two Images |
| 8.3.3 Projective Structure and Motion from Multiple Images | | 8.3.3 Projective Structure and Motion from Multiple Images |
| 8.3.4 From Projective to Euclidean Shape | | 8.3.4 From Projective to Euclidean Shape |

| I | V N | MID-LEVEL VISION | 283 |
|----|-------|---|------------|
| 9 | Seg | mentation by Clustering | 285 |
| | 9.1 | Human Vision: Grouping and Gestalt | 286 |
| | 9.2 | Important Applications | 291 |
| | | 9.2.1 Background Subtraction | 291 |
| | | 9.2.2 Shot Boundary Detection | 294 |
| | | 9.2.3 Interactive Segmentation | 295 |
| | | 9.2.4 Forming Image Regions | 296 |
| | 9.3 | Image Segmentation by Clustering Pixels | 298 |
| | | 9.3.1 Basic Clustering Methods | 299 |
| | | 9.3.2 The Watershed Algorithm | 301 |
| | | 9.3.3 Segmentation Using K-means | 302 |
| | | 9.3.4 Mean Shift: Finding Local Modes in Data | 303 |
| | | 9.3.5 Clustering and Segmentation with Mean Shift | 305 |
| | 9.4 | Segmentation, Clustering, and Graphs | 307 |
| | | 9.4.1 Terminology and Facts for Graphs | 307 |
| | | 9.4.2 Agglomerative Clustering with a Graph | 309 |
| | | 9.4.3 Divisive Clustering with a Graph | 311 |
| | | 9.4.4 Normalized Cuts | 314 |
| | 9.5 | Image Segmentation in Practice | 315 |
| | | 9.5.1 Evaluating Segmenters | 316 |
| | 9.6 | Notes | 317 |
| 10 |) Gro | ouping and Model Fitting | 320 |
| | 10.1 | The Hough Transform | 320 |
| | | 10.1.1 Fitting Lines with the Hough Transform | 320 |
| | | 10.1.2 Using the Hough Transform | 322 |
| | 10.2 | Fitting Lines and Planes | 323 |
| | | 10.2.1 Fitting a Single Line | 324 |
| | | 10.2.2 Fitting Planes | 325 |
| | | 10.2.3 Fitting Multiple Lines | 326 |
| | 10.3 | Fitting Curved Structures | 327 |
| | 10.4 | Robustness | 329 |
| | | 10.4.1 M-Estimators | 330 |
| | | 10.4.2 RANSAC: Searching for Good Points | 332 |
| | 10.5 | Fitting Using Probabilistic Models | 336 |
| | | 10.5.1 Missing Data Problems | 337 |
| | | 10.5.2 Mixture Models and Hidden Variables | 339 |
| | | 10.5.3 The EM Algorithm for Mixture Models | 340 |
| | | 10.5.4 Difficulties with the EM Algorithm | 342 |

| | 10.6 | Motion Segmentation by Parameter Estimation | 343 |
|------------|------|---|-----|
| | | 10.6.1 Optical Flow and Motion | 345 |
| | | 10.6.2 Flow Models | 346 |
| | | 10.6.3 Motion Segmentation with Layers | 347 |
| | 10.7 | Model Selection: Which Model Is the Best Fit? | 349 |
| | | 10.7.1 Model Selection Using Cross-Validation | 352 |
| | 10.8 | Notes | 352 |
| 11 | Trac | king 3 | 56 |
| | 11.1 | Simple Tracking Strategies | 357 |
| | | 11.1.1 Tracking by Detection | 357 |
| | | 11.1.2 Tracking Translations by Matching | 360 |
| | | 11.1.3 Using Affine Transformations to Confirm a Match | 362 |
| | 11.2 | Tracking Using Matching | 364 |
| | | 11.2.1 Matching Summary Representations | 365 |
| | | 11.2.2 Tracking Using Flow | 367 |
| | 11.3 | Tracking Linear Dynamical Models with Kalman Filters | 369 |
| | | 11.3.1 Linear Measurements and Linear Dynamics | 370 |
| | | 11.3.2 The Kalman Filter | 374 |
| | | 11.3.3 Forward-backward Smoothing | 375 |
| | 11.4 | Data Association | 379 |
| | | 11.4.1 Linking Kalman Filters with Detection Methods | 379 |
| | | 11.4.2 Key Methods of Data Association | 380 |
| | 11.5 | Particle Filtering | 380 |
| | | 11.5.1 Sampled Representations of Probability Distributions | 381 |
| | | 11.5.2 The Simplest Particle Filter | 385 |
| | | 11.5.3 The Tracking Algorithm | 386 |
| | | 11.5.4 A Workable Particle Filter | 388 |
| | | 11.5.5 Practical Issues in Particle Filters | 390 |
| | 11.6 | Notes | 392 |
| 1 7 | 17 | | ٥٢ |
| v | Н | IGH-LEVEL VISION 3 | 95 |
| 12 | Reg | stration 3 | 897 |
| | 12.1 | Registering Rigid Objects | 398 |
| | | 12.1.1 Iterated Closest Points | 398 |

| | 12.1.1 | |
|------|--------|--|
| | 12.1.2 | Searching for Transformations via Correspondences 399 |
| | 12.1.3 | Application: Building Image Mosaics |
| 12.2 | Model | -based Vision: Registering Rigid Objects with Projection 405 |
| | | |

| 12.2.1 Verification: Comparing Transformed and Rendered Source |
|---|
| to Target $\ldots \ldots 407$ |
| 12.3 Registering Deformable Objects |
| 12.3.1 Deforming Texture with Active Appearance Models 408 |
| 12.3.2 Active Appearance Models in Practice |
| 12.3.3 Application: Registration in Medical Imaging Systems 413 |
| 12.4 Notes |
| 13 Smooth Surfaces and Their Outlines 421 |
| 13.1 Elements of Differential Geometry |
| 13.1.1 Curves |
| 13.1.2 Surfaces |
| 13.2 Contour Geometry |
| 13.2.1 The Occluding Contour and the Image Contour |
| 13.2.2 The Cusps and Inflections of the Image Contour |
| 13.2.3 Koenderink's Theorem |
| 13.3 Visual Events: More Differential Geometry |
| 13.3.1 The Geometry of the Gauss Map $\ldots \ldots \ldots \ldots \ldots 437$ |
| 13.3.2 Asymptotic Curves |
| 13.3.3 The Asymptotic Spherical Map |
| 13.3.4 Local Visual Events |
| 13.3.5 The Bitangent Ray Manifold |
| 13.3.6 Multilocal Visual Events |
| 13.3.7 The Aspect Graph |
| 13.4 Notes |
| 14 Range Data 452 |
| 14.1 Active Range Sensors |
| 14.2 Range Data Segmentation |
| 14.2.1 Elements of Analytical Differential Geometry |
| 14.2.2 Finding Step and Roof Edges in Range Images |
| 14.2.3 Segmenting Range Images into Planar Regions |
| 14.3 Range Image Registration and Model Acquisition |
| 14.3.1 Quaternions |
| 14.3.2 Registering Range Images |

14.4 Object Recognition46814.4.1 Matching Using Interpretation Trees46814.4.2 Matching Free-Form Surfaces Using Spin Images47114.5 Kinect47614.5.1 Features477

| | 14. | 5.2 | Technique: Decision Trees and Random Forests |
|-------|---------|----------------------|---|
| | 14. | 5.3 | Labeling Pixels |
| | 14. | 5.4 | Computing Joint Positions |
| 14 | l.6 No | tes | |
| 15 Le | earnin | ıg t | o Classify 487 |
| 15 | 5.1 Cla | assif | ication, Error, and Loss |
| | 15. | 1.1 | Using Loss to Determine Decisions |
| | 15. | 1.2 | Training Error, Test Error, and Overfitting |
| | 15. | 1.3 | Regularization |
| | 15. | 1.4 | Error Rate and Cross-Validation |
| | 15. | 1.5 | Receiver Operating Curves |
| 15 | 5.2 Ma | jor | Classification Strategies |
| | 15. | 2.1 | Example: Mahalanobis Distance |
| | 15. | 2.2 | Example: Class-Conditional Histograms and Naive Bayes $\ . \ . \ 498$ |
| | 15. | 2.3 | Example: Classification Using Nearest Neighbors 499 |
| | 15. | 2.4 | Example: The Linear Support Vector Machine |
| | 15. | 2.5 | Example: Kernel Machines |
| | 15. | 2.6 | Example: Boosting and Adaboost |
| 15 | 5.3 Pra | actic | cal Methods for Building Classifiers |
| | 15. | 3.1 | Manipulating Training Data to Improve Performance 507 |
| | 15. | 3.2 | Building Multi-Class Classifiers Out of Binary Classifiers $~$ 509 |
| | 15. | 3.3 | Solving for SVMS and Kernel Machines |
| 15 | 6.4 No | tes | |
| 16 C | lassify | ving | g Images 512 |
| 16 | .1 Bu | ildir | ng Good Image Features |
| | 16. | 1.1 | Example Applications |
| | 16. | 1.2 | Encoding Layout with GIST Features |
| | 16. | 1.3 | Summarizing Images with Visual Words $\ldots \ldots \ldots \ldots \ldots 517$ |
| | 16. | 1.4 | The Spatial Pyramid Kernel |
| | 16. | 1.5 | Dimension Reduction with Principal Components |
| | 16. | 1.6 | Dimension Reduction with Canonical Variates $\ldots \ldots \ldots 524$ |
| | 16. | 1.7 | Example Application: Identifying Explicit Images 528 |
| | 16. | 1.8 | Example Application: Classifying Materials |
| | 16. | 1.9 | Example Application: Classifying Scenes |
| 16 | 5.2 Cla | assif | ying Images of Single Objects |
| | 16. | 2.1 | Image Classification Strategies |
| | 16. | 2.2 | Evaluating Image Classification Systems |

| 16.2.5 Flowers, Leaves, and Birds: Some Specialized Problems54116.3 Image Classification in Practice54216.3.1 Codes for Image Features54316.3.2 Image Classification Datasets54316.3.3 Dataset Bias54516.3.4 Crowdsourcing Dataset Collection545 |
|--|
| 16.4 Notes |
| 17 Detecting Objects in Images 549 |
| 17.1 The Sliding Window Method |
| 17.1.1 Face Detection |
| 17.1.2 Detecting Humans |
| 17.1.3 Detecting Boundaries |
| 17.2 Detecting Deformable Objects |
| 17.3 The State of the Art of Object Detection |
| 17.3.1 Datasets and Resources |
| 17.4 Notes |
| 18 Topics in Object Recognition 570 |
| 18.1 What Should Object Recognition Do? |
| 18.1.1 What Should an Object Recognition System Do? |
| 18.1.2 Current Strategies for Object Recognition |
| 18.1.3 What Is Categorization? |
| 18.1.4 Selection: What Should Be Described? |
| 18.2 Feature Questions |
| 18.2.1 Improving Current Image Features |
| 18.2.2 Other Kinds of Image Feature |
| 18.3 Geometric Questions |
| 18.4 Semantic Questions |
| 18.4.1 Attributes and the Unfamiliar |
| 18.4.2 Parts, Poselets and Consistency |
| 18.4.3 Chunks of Meaning |
| VI APPLICATIONS AND TODICS 597 |

19 Image-Based Modeling and Rendering58919.1 Visual Hulls58919.1.1 Main Elements of the Visual Hull Model59119.1.2 Tracing Intersection Curves59319.1.3 Clipping Intersection Curves596

| 19.1.4 | Triangulating Cone Strips |
|-------------|------------------------------------|
| 19.1.5 | Results |
| 19.1.6 | Going Further: Carved Visual Hulls |
| 19.2 Patch- | Based Multi-View Stereopsis |
| 19.2.1 | Main Elements of the PMVS Model |
| 19.2.2 | Initial Feature Matching |
| 19.2.3 | Expansion |
| 19.2.4 | Filtering |
| 19.2.5 | Results |
| 19.3 The L | ight Field |
| 19.4 Notes | |

| 20 | 20 Looking at People620 | | | |
|--|-------------------------|---|--|-------------|
| | 20.1 | 1 HMM's, Dynamic Programming, and Tree-Structured Models 62 | | |
| | | 20.1.1 | Hidden Markov Models | 520 |
| | | 20.1.2 | Inference for an HMM | 522 |
| | | 20.1.3 | Fitting an HMM with EM | 527 |
| | | 20.1.4 | Tree-Structured Energy Models | i 30 |
| | 20.2 | .2 Parsing People in Images | | 532 |
| | | 20.2.1 | Parsing with Pictorial Structure Models 6 | 532 |
| | | 20.2.2 | Estimating the Appearance of Clothing | 534 |
| | 20.3 | Tracki | ng People | 536 |
| | | 20.3.1 | Why Human Tracking Is Hard | 536 |
| | | 20.3.2 | Kinematic Tracking by Appearance 6 | 538 |
| 20.3.3 Kinematic Human Tracking Using Templa | | 20.3.3 | Kinematic Human Tracking Using Templates | 539 |
| | 20.4 | 3D fro | m 2D: Lifting | 541 |
| | | 20.4.1 | Reconstruction in an Orthographic View 6 | 541 |
| | | 20.4.2 | Exploiting Appearance for Unambiguous Reconstructions 6 | 543 |
| | | 20.4.3 | Exploiting Motion for Unambiguous Reconstructions 6 | 645 |
| | 20.5 | 20.5 Activity Recognition | | 547 |
| | | 20.5.1 | Background: Human Motion Data | 547 |
| | | 20.5.2 | Body Configuration and Activity Recognition 6 | 551 |
| | | 20.5.3 | Recognizing Human Activities with Appearance Features 6 | 652 |
| | | 20.5.4 | Recognizing Human Activities with Compositional Models 6 | 354 |
| | 20.6 | Resour | rces | 354 |
| | 20.7 | Notes | | 556 |
| 21 | Ima | ge Sea | rch and Retrieval 6 | 57 |
| | 21.1 | 21.1 The Application Context | | |
| | | 21.1.1 | Applications | 358 |

| | 21.1.4 | What Users Do with Image Collections $\hfill \ldots \ldots \ldots \ldots \ldots 661$ |
|------|---------|--|
| 21.2 | Basic 7 | Fechnologies from Information Retrieval 662 |
| | 21.2.1 | Word Counts |
| | 21.2.2 | Smoothing Word Counts |
| 21.3 | 21.2.3 | Approximate Nearest Neighbors and Hashing |
| | 21.2.4 | Ranking Documents |
| | Images | as Documents |
| | 21.3.1 | Matching Without Quantization |
| | 21.3.2 | Ranking Image Search Results |

| | 21.3.1 | Matching Without Quantization |
|------|--------|--|
| | 21.3.2 | Ranking Image Search Results |
| | 21.3.3 | Browsing and Layout |
| | 21.3.4 | Laying Out Images for Browsing |
| 21.4 | Predic | ting Annotations for Pictures |
| | 21.4.1 | Annotations from Nearby Words |
| | 21.4.2 | Annotations from the Whole Image |
| | 21.4.3 | Predicting Correlated Words with Classifiers |
| | 21.4.4 | Names and Faces |
| | 21.4.5 | Generating Tags with Segments |
| 21.5 | The St | ate of the Art of Word Prediction |
| | 21.5.1 | Resources |
| | 21.5.2 | Comparing Methods |
| | 21.5.3 | Open Problems |
| 21.6 | Notes | |

VII BACKGROUND MATERIAL

| 22 Optimization Techniques 6 | | | 69 | 93 |
|--------------------------------------|-----------------------------------|--|----|----|
| 22.1 | 22.1 Linear Least-Squares Methods | | 6 | 93 |
| | 22.1.1 | Normal Equations and the Pseudoinverse | 6 | 94 |
| | 22.1.2 | Homogeneous Systems and Eigenvalue Problems | 6 | 95 |
| | 22.1.3 | Generalized Eigenvalues Problems | 6 | 96 |
| | 22.1.4 | An Example: Fitting a Line to Points in a Plane | 69 | 96 |
| | 22.1.5 | Singular Value Decomposition | 6 | 97 |
| 22.2 Nonlinear Least-Squares Methods | | 6 | 99 | |
| | 22.2.1 | Newton's Method: Square Systems of Nonlinear Equations | 7 | 00 |
| | 22.2.2 | Newton's Method for Overconstrained Systems | 70 | 00 |
| | 22.2.3 | The Gauss–Newton and Levenberg–Marquardt Algorithms . | 7 | 01 |
| 22.3 | Sparse | Coding and Dictionary Learning | 7 | 02 |
| | 22.3.1 | Sparse Coding | 7 | 02 |
| | 22.3.2 | Dictionary Learning | 7 | 03 |
| | | | | |

| 22.3.3 | Supervised Dictionary Learning | . 705 |
|-----------------|--|-------|
| 22.4 Min-Cu | t/Max-Flow Problems and Combinatorial Optimization $\ . \ .$ | . 705 |
| 22.4.1 | Min-Cut Problems | . 706 |
| 22.4.2 | Quadratic Pseudo-Boolean Functions | . 707 |
| 22.4.3 | Generalization to Integer Variables | . 709 |
| 22.5 Notes | | . 712 |
| Bibliography | | 714 |
| Index | | 767 |
| List of Algorit | hms | 790 |