## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>xxi</td>
</tr>
<tr>
<td>Notation</td>
<td>xxvii</td>
</tr>
<tr>
<td><strong>1 The Challenge of Vision</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Vision</td>
<td>1</td>
</tr>
<tr>
<td>1.3 Theories of Vision</td>
<td>4</td>
</tr>
<tr>
<td>1.4 What’s Next?</td>
<td>31</td>
</tr>
<tr>
<td>1.5 Concluding Remarks</td>
<td>32</td>
</tr>
<tr>
<td><strong>1 FOUNDATIONS</strong></td>
<td>33</td>
</tr>
<tr>
<td><strong>2 A Simple Vision System</strong></td>
<td>35</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>35</td>
</tr>
<tr>
<td>2.2 A Simple World: The Blocks World</td>
<td>35</td>
</tr>
<tr>
<td>2.3 A Simple Image Formation Model</td>
<td>36</td>
</tr>
<tr>
<td>2.4 A Simple Goal</td>
<td>38</td>
</tr>
<tr>
<td>2.5 From Images to Edges and Useful Features</td>
<td>38</td>
</tr>
<tr>
<td>2.6 From Edges to Surfaces</td>
<td>42</td>
</tr>
<tr>
<td>2.7 Generalization</td>
<td>49</td>
</tr>
<tr>
<td>2.8 Concluding Remarks</td>
<td>51</td>
</tr>
<tr>
<td><strong>3 Looking at Images</strong></td>
<td>53</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>53</td>
</tr>
<tr>
<td>3.2 Looking at Individual Pixels</td>
<td>53</td>
</tr>
<tr>
<td>3.3 The More You Look, the More You See</td>
<td>55</td>
</tr>
<tr>
<td>3.4 The Eye of the Artist</td>
<td>57</td>
</tr>
<tr>
<td>3.5 Tree Shadows and Image Formation</td>
<td>58</td>
</tr>
<tr>
<td>3.6 Horizontal or Vertical</td>
<td>59</td>
</tr>
<tr>
<td>3.7 Motion Blur</td>
<td>60</td>
</tr>
<tr>
<td>3.8 Accidents Happen</td>
<td>62</td>
</tr>
<tr>
<td>3.9 Cues for Support</td>
<td>63</td>
</tr>
<tr>
<td>3.10 Looking at Raindrops</td>
<td>64</td>
</tr>
<tr>
<td>3.11 Plato’s Cave</td>
<td>65</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>3.12</td>
<td>How Do You Know Something Is Wet?</td>
</tr>
<tr>
<td>3.13</td>
<td>Concluding Remarks</td>
</tr>
<tr>
<td>4</td>
<td>Computer Vision and Society</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>4.2</td>
<td>Fairness</td>
</tr>
<tr>
<td>4.3</td>
<td>Ethics</td>
</tr>
<tr>
<td>4.4</td>
<td>Concluding Remarks</td>
</tr>
<tr>
<td>II</td>
<td>IMAGE FORMATION</td>
</tr>
<tr>
<td>5</td>
<td>Imaging</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>5.2</td>
<td>Light Interacting with Surfaces</td>
</tr>
<tr>
<td>5.3</td>
<td>The Pinhole Camera and Image Formation</td>
</tr>
<tr>
<td>5.4</td>
<td>Concluding Remarks</td>
</tr>
<tr>
<td>6</td>
<td>Lenses</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>6.2</td>
<td>Lensmaker's Formula</td>
</tr>
<tr>
<td>6.3</td>
<td>Imaging with Lenses</td>
</tr>
<tr>
<td>6.4</td>
<td>Concluding Remarks</td>
</tr>
<tr>
<td>7</td>
<td>Cameras as Linear Systems</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>7.2</td>
<td>Flatland</td>
</tr>
<tr>
<td>7.3</td>
<td>Cameras as Linear Systems</td>
</tr>
<tr>
<td>7.4</td>
<td>More General Imagers</td>
</tr>
<tr>
<td>7.5</td>
<td>Concluding Remarks</td>
</tr>
<tr>
<td>8</td>
<td>Color</td>
</tr>
<tr>
<td>8.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>8.2</td>
<td>Color Physics</td>
</tr>
<tr>
<td>8.3</td>
<td>Color Perception</td>
</tr>
<tr>
<td>8.4</td>
<td>Spatial Resolution and Color</td>
</tr>
<tr>
<td>8.5</td>
<td>Concluding Remarks</td>
</tr>
<tr>
<td>III</td>
<td>FOUNDATIONS OF LEARNING</td>
</tr>
<tr>
<td>9</td>
<td>Introduction to Learning</td>
</tr>
<tr>
<td>9.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>9.2</td>
<td>Learning from Examples</td>
</tr>
<tr>
<td>9.3</td>
<td>Learning without Examples</td>
</tr>
<tr>
<td>9.4</td>
<td>Key Ingredients</td>
</tr>
</tbody>
</table>
9.5 Empirical Risk Minimization: A Formalization of Learning from Examples 141
9.6 Learning as Probabilistic Inference 142
9.7 Case Studies 142
9.8 Learning to Learn 149
9.9 Concluding Remarks 150

10 Gradient-Based Learning Algorithms 151
10.1 Introduction 151
10.2 Technical Setting 151
10.3 Basic Gradient Descent 152
10.4 Learning Rate Schedules 153
10.5 Momentum 153
10.6 What Kinds of Functions Can Be Minimized with Gradient Descent? 155
10.7 Stochastic Gradient Descent 159
10.8 Concluding Remarks 160

11 The Problem of Generalization 161
11.1 Introduction 161
11.2 Underfitting and Overfitting 161
11.3 Regularization 165
11.4 Rethinking Generalization 167
11.5 Three Tools in the Search for Truth: Data, Priors, and Hypotheses 167
11.6 Concluding Remarks 173

12 Neural Networks 175
12.1 Introduction 175
12.2 The Perceptron: A Simple Model of a Single Neuron 175
12.3 Multilayer Perceptrons 177
12.4 Activations Versus Parameters 179
12.5 Deep Nets 180
12.6 Deep Learning: Learning with Neural Nets 184
12.7 Catalog of Layers 186
12.8 Why Are Neural Networks a Good Architecture? 189
12.9 Concluding Remarks 190

13 Neural Networks as Distribution Transformers 191
13.1 Introduction 191
13.2 A Different Way of Plotting Functions 191
13.3 How Deep Nets Remap a Data Distribution 193
13.4 Binary Classifier Example 194
13.5 How High-Dimensional Datapoints Get Remapped by Deep Net 196
13.6 Concluding Remarks 198
## Backpropagation

14.1 Introduction .............................................................. 199
14.2 The Trick of Backpropagation: Reuse of Computation ................ 200
14.3 Backward for a Generic Layer ....................................... 201
14.4 The Full Algorithm: Forward, Then Backward ...................... 203
14.5 Backpropagation Over Data Batches .................................. 204
14.6 Example: Backpropagation for an MLP .............................. 205
14.7 Backpropagation through DAGs: Branch and Merge ............... 212
14.8 Parameter Sharing ..................................................... 214
14.9 Backpropagation to the Data ......................................... 214
14.10 Concluding Remarks ................................................... 216

### IV FOUNDATIONS OF IMAGE PROCESSING

15 Linear Image Filtering ..................................................... 219
15.1 Introduction .................................................................. 219
15.2 Signals and Images ....................................................... 219
15.3 Systems ..................................................................... 223
15.4 Convolution .................................................................. 227
15.5 Cross-Correlation Versus Convolution ............................... 235
15.6 System Identification ..................................................... 238
15.7 Concluding Remarks ..................................................... 239

16 Fourier Analysis ............................................................... 241
16.1 Introduction .................................................................. 241
16.2 Image Transforms ........................................................ 241
16.3 Fourier Series .............................................................. 241
16.4 Continuous and Discrete Waves ....................................... 243
16.5 The Discrete Fourier Transform ....................................... 247
16.6 Useful Transforms ........................................................ 251
16.7 Discrete Fourier Transform Properties .............................. 255
16.8 A Family of Fourier Transforms ...................................... 261
16.9 Fourier Analysis as an Image Representation ...................... 262
16.10 Fourier Analysis of Linear Filters ................................... 267
16.11 Concluding Remarks ................................................... 272

### V LINEAR FILTERS

17 Blur Filters ................................................................. 275
17.1 Introduction .................................................................. 275
17.2 Box Filter ................................................................... 276
17.3 Gaussian Filter ............................................................ 279
17.4 Binomial Filters ........................................................... 283
## Contents

17.5 Concluding Remarks ................................................................. 286

18 **Image Derivatives** ................................................................. 287
  18.1 Introduction ................................................................................. 287
  18.2 Discretizing Image Derivatives .................................................. 287
  18.3 Gradient-Based Image Representation ........................................ 291
  18.4 Image Editing in the Gradient Domain ....................................... 292
  18.5 Gaussian Derivatives ................................................................. 293
  18.6 High-Order Gaussian Derivatives .............................................. 295
  18.7 Derivatives of Binomial Filters ................................................. 299
  18.8 Image Gradient and Directional Derivatives .................................. 301
  18.9 Image Laplacian ........................................................................ 302
  18.10 A Simple Model of the Early Visual System .............................. 305
  18.11 Sharpening Filter ..................................................................... 307
  18.12 Retinex .................................................................................... 309
  18.13 Concluding Remarks ............................................................... 313

19 **Temporal Filters** ........................................................................ 315
  19.1 Introduction ................................................................................. 315
  19.2 Modeling Sequences ................................................................... 315
  19.3 Modeling Sequences in the Fourier Domain ............................... 317
  19.4 Temporal Filters ........................................................................ 318
  19.5 Concluding Remarks ............................................................... 324

VI **SAMPLING AND MULTISCALE IMAGE REPRESENTATIONS** ........ 325

20 **Image Sampling and Aliasing** .................................................. 327
  20.1 Introduction ................................................................................. 327
  20.2 Aliasing ....................................................................................... 327
  20.3 Sampling Theorem ...................................................................... 329
  20.4 Reconstruction ............................................................................ 334
  20.5 Ideal Reconstruction .................................................................... 334
  20.6 A Family of 2D Spatial Samplings ............................................. 338
  20.7 Anti-Aliasing Filter ..................................................................... 340
  20.8 Spatiotemporal Sampling ........................................................... 342
  20.9 Concluding Remarks ............................................................... 342

21 **Downsampling and Upsampling Images** .................................... 345
  21.1 Introduction ................................................................................. 345
  21.2 Example: Aliasing-Based Adversarial Attack .............................. 345
  21.3 Downsampling ........................................................................... 346
  21.4 Upsampling ............................................................................... 358
  21.5 Concluding Remarks ............................................................... 363
## Contents

**22 Filter Banks**
- 22.1 Introduction ............................................................... 365
- 22.2 Gabor Filters ............................................................. 365
- 22.3 Steerable Filters and Orientation Analysis ......................... 374
- 22.4 Motion Analysis .......................................................... 380
- 22.5 Concluding Remarks .................................................... 383

**23 Image Pyramids**
- 23.1 Introduction ............................................................. 385
- 23.2 Image Pyramids and Multiscale Image Analysis ...................... 386
- 23.3 Linear Image Transforms ................................................. 387
- 23.4 Gaussian Pyramid ........................................................ 388
- 23.5 Laplacian Pyramid ....................................................... 390
- 23.6 Steerable Pyramid ....................................................... 395
- 23.7 A Pictorial Summary ..................................................... 397
- 23.8 Concluding Remarks ..................................................... 399

**VII NEURAL ARCHITECTURES FOR VISION** ........................................ 401

**24 Convolutional Neural Nets** .................................................. 403
- 24.1 Introduction ............................................................. 403
- 24.2 Convolutional Layers ................................................... 404
- 24.3 Nonlinear Filtering Layers .............................................. 414
- 24.4 A Simple CNN Classifier .............................................. 415
- 24.5 A Worked Example ..................................................... 417
- 24.6 Feature Maps in CNNs ................................................... 420
- 24.7 Receptive Fields ........................................................ 423
- 24.8 Spatial Outputs .......................................................... 424
- 24.9 CNN as a Sliding Filter ................................................ 425
- 24.11 Popular CNN Architectures ........................................... 427
- 24.12 Concluding Remarks ................................................... 430

**25 Recurrent Neural Nets** ...................................................... 431
- 25.1 Introduction ............................................................. 431
- 25.2 Recurrent Layer .......................................................... 433
- 25.3 Backpropagation through Time ........................................ 433
- 25.4 Stacking Recurrent Layers .............................................. 435
- 25.5 Long Short-Term Memory .............................................. 436
- 25.6 Concluding Remarks .................................................... 437

**26 Transformers** ............................................................... 439
- 26.1 Introduction ............................................................. 439
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>26.2</td>
<td>A Limitation of CNNs: Independence between Far Apart Patches</td>
<td>439</td>
</tr>
<tr>
<td>26.3</td>
<td>The Idea of Attention</td>
<td>440</td>
</tr>
<tr>
<td>26.4</td>
<td>A New Data Type: Tokens</td>
<td>440</td>
</tr>
<tr>
<td>26.5</td>
<td>Token Nets</td>
<td>444</td>
</tr>
<tr>
<td>26.6</td>
<td>The Attention Layer</td>
<td>445</td>
</tr>
<tr>
<td>26.7</td>
<td>The Full Transformer Architecture</td>
<td>453</td>
</tr>
<tr>
<td>26.8</td>
<td>Permutation Equivariance</td>
<td>455</td>
</tr>
<tr>
<td>26.9</td>
<td>CNNs in Disguise</td>
<td>456</td>
</tr>
<tr>
<td>26.10</td>
<td>Masked Attention</td>
<td>458</td>
</tr>
<tr>
<td>26.11</td>
<td>Positional Encodings</td>
<td>460</td>
</tr>
<tr>
<td>26.12</td>
<td>Comparing Fully Connected, Convolutional, and Self-Attention Layers</td>
<td>462</td>
</tr>
<tr>
<td>26.13</td>
<td>Concluding Remarks</td>
<td>463</td>
</tr>
</tbody>
</table>

**VIII PROBABILISTIC MODELS OF IMAGES**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>Statistical Image Models</td>
<td>465</td>
</tr>
<tr>
<td>27.1</td>
<td>Introduction</td>
<td>467</td>
</tr>
<tr>
<td>27.2</td>
<td>How Do We Tell Noise from Texture?</td>
<td>469</td>
</tr>
<tr>
<td>27.3</td>
<td>Independent Pixels</td>
<td>470</td>
</tr>
<tr>
<td>27.4</td>
<td>Dead Leaves Model</td>
<td>474</td>
</tr>
<tr>
<td>27.5</td>
<td>The Gaussian Model</td>
<td>477</td>
</tr>
<tr>
<td>27.6</td>
<td>The Wavelet Marginal Model</td>
<td>482</td>
</tr>
<tr>
<td>27.7</td>
<td>Nonparametric Markov Random Field Image Models</td>
<td>489</td>
</tr>
<tr>
<td>27.8</td>
<td>Concluding Remarks</td>
<td>490</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>Textures</td>
<td>493</td>
</tr>
<tr>
<td>28.1</td>
<td>Introduction</td>
<td>493</td>
</tr>
<tr>
<td>28.2</td>
<td>A Few Notes about Human Perception</td>
<td>494</td>
</tr>
<tr>
<td>28.3</td>
<td>Heeger-Bergen Texture Analysis and Synthesis</td>
<td>496</td>
</tr>
<tr>
<td>28.4</td>
<td>Efros-Leung Texture Analysis and Synthesis Model</td>
<td>501</td>
</tr>
<tr>
<td>28.5</td>
<td>Connection to Deep Generative Models</td>
<td>503</td>
</tr>
<tr>
<td>28.6</td>
<td>Concluding Remarks</td>
<td>504</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>Probabilistic Graphical Models</td>
<td>505</td>
</tr>
<tr>
<td>29.1</td>
<td>Introduction</td>
<td>505</td>
</tr>
<tr>
<td>29.2</td>
<td>Simple Examples</td>
<td>505</td>
</tr>
<tr>
<td>29.3</td>
<td>Directed Graphical Models</td>
<td>509</td>
</tr>
<tr>
<td>29.4</td>
<td>Inference in Graphical Models</td>
<td>510</td>
</tr>
<tr>
<td>29.5</td>
<td>Simple Example of Inference in a Graphical Model</td>
<td>511</td>
</tr>
<tr>
<td>29.6</td>
<td>Belief Propagation</td>
<td>512</td>
</tr>
<tr>
<td>29.7</td>
<td>Loopy Belief Propagation</td>
<td>520</td>
</tr>
<tr>
<td>29.8</td>
<td>Relationship of Probabilistic Graphical Models to Neural Networks</td>
<td>523</td>
</tr>
<tr>
<td>29.9</td>
<td>Concluding Remarks</td>
<td>523</td>
</tr>
</tbody>
</table>
IX GENERATIVE IMAGE MODELS AND REPRESENTATION LEARNING .......... 525

30 Representation Learning ............................................. 527
  30.1 Introduction ................................................... 527
  30.2 Problem Setting ............................................... 527
  30.3 What Makes for a Good Representation? ....................... 528
  30.4 Autoencoders ............................................... 530
  30.5 Predictive Encodings ......................................... 533
  30.6 Self-Supervised Learning ................................... 535
  30.7 Imputation ................................................. 536
  30.8 Abstract Pretext Tasks ....................................... 537
  30.9 Clustering .................................................. 537
  30.10 Contrastive Learning ....................................... 542
  30.11 Concluding Remarks ........................................ 547

31 Perceptual Grouping .................................................. 549
  31.1 Introduction ................................................... 549
  31.2 Why Group? .................................................. 550
  31.3 Segments ..................................................... 551
  31.4 Edges, Boundaries, and Contours ............................ 555
  31.5 Layers ........................................................ 556
  31.6 Emergent Groups ............................................. 556
  31.7 Concluding Remarks ........................................ 557

32 Generative Models .................................................... 559
  32.1 Introduction ................................................... 559
  32.2 Unconditional Generative Models ............................. 561
  32.3 Learning Generative Models ................................ 563
  32.4 Density Models ............................................... 565
  32.5 Energy-Based Models ........................................ 567
  32.6 Gaussian Density Models .................................... 570
  32.7 Autoregressive Density Models ................................ 572
  32.8 Diffusion Models ............................................. 576
  32.9 Generative Adversarial Networks ............................ 579
  32.10 Concluding Remarks ......................................... 581

33 Generative Modeling Meets Representation Learning ............... 583
  33.1 Introduction ................................................... 583
  33.2 Latent Variables as Representations .......................... 584
  33.3 Technical Setting ............................................. 585
  33.4 Variational Autoencoders .................................... 586
  33.5 Do VAEs Learn Good Representations? ....................... 598
  33.6 Generative Adversarial Networks Are Representation Learners Too .......................................................... 600
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>33.7</td>
<td>Concluding Remarks</td>
<td>601</td>
</tr>
<tr>
<td>34</td>
<td><strong>Conditional Generative Models</strong></td>
<td>603</td>
</tr>
<tr>
<td>34.1</td>
<td>Introduction</td>
<td>603</td>
</tr>
<tr>
<td>34.2</td>
<td>A Motivating Example: Image Colorization</td>
<td>603</td>
</tr>
<tr>
<td>34.3</td>
<td>Conditional Generative Models Solve Multimodal Structured Prediction</td>
<td>608</td>
</tr>
<tr>
<td>34.4</td>
<td>A Tour of Popular Conditional Models</td>
<td>609</td>
</tr>
<tr>
<td>34.5</td>
<td>Structured Prediction in Vision</td>
<td>613</td>
</tr>
<tr>
<td>34.6</td>
<td>Image-to-Image Translation</td>
<td>614</td>
</tr>
<tr>
<td>34.7</td>
<td>Concluding Remarks</td>
<td>620</td>
</tr>
<tr>
<td>X</td>
<td><strong>CHALLENGES IN LEARNING-BASED VISION</strong></td>
<td>621</td>
</tr>
<tr>
<td>35</td>
<td><strong>Data Bias and Shift</strong></td>
<td>623</td>
</tr>
<tr>
<td>35.1</td>
<td>Introduction</td>
<td>623</td>
</tr>
<tr>
<td>35.2</td>
<td>Out-of-Distribution Generalization</td>
<td>625</td>
</tr>
<tr>
<td>35.3</td>
<td>A Toy Example</td>
<td>627</td>
</tr>
<tr>
<td>35.4</td>
<td>Dataset Bias</td>
<td>630</td>
</tr>
<tr>
<td>35.5</td>
<td>Sources of Bias</td>
<td>632</td>
</tr>
<tr>
<td>35.6</td>
<td>Adversarial Shifts</td>
<td>636</td>
</tr>
<tr>
<td>35.7</td>
<td>Concluding Remarks</td>
<td>637</td>
</tr>
<tr>
<td>36</td>
<td><strong>Training for Robustness and Generality</strong></td>
<td>639</td>
</tr>
<tr>
<td>36.1</td>
<td>Introduction</td>
<td>639</td>
</tr>
<tr>
<td>36.2</td>
<td>Data Augmentation</td>
<td>639</td>
</tr>
<tr>
<td>36.3</td>
<td>Adversarial Training</td>
<td>643</td>
</tr>
<tr>
<td>36.4</td>
<td>Toward General-Purpose Vision Models</td>
<td>643</td>
</tr>
<tr>
<td>36.5</td>
<td>Concluding Remarks</td>
<td>644</td>
</tr>
<tr>
<td>37</td>
<td><strong>Transfer Learning and Adaptation</strong></td>
<td>645</td>
</tr>
<tr>
<td>37.1</td>
<td>Introduction</td>
<td>645</td>
</tr>
<tr>
<td>37.2</td>
<td>Problem Setting</td>
<td>645</td>
</tr>
<tr>
<td>37.3</td>
<td>Finetuning</td>
<td>646</td>
</tr>
<tr>
<td>37.4</td>
<td>Learning from a Teacher</td>
<td>649</td>
</tr>
<tr>
<td>37.5</td>
<td>Prompting</td>
<td>651</td>
</tr>
<tr>
<td>37.6</td>
<td>Domain Adaptation</td>
<td>653</td>
</tr>
<tr>
<td>37.7</td>
<td>Generative Data</td>
<td>654</td>
</tr>
<tr>
<td>37.8</td>
<td>Other Kinds of Knowledge that Can Be Transferred</td>
<td>655</td>
</tr>
<tr>
<td>37.9</td>
<td>A Combinatorial Catalog of Transfer Learning Methods</td>
<td>655</td>
</tr>
<tr>
<td>37.10</td>
<td>Sequence Models from the Lens of Adaptation</td>
<td>656</td>
</tr>
<tr>
<td>37.11</td>
<td>Concluding Remarks</td>
<td>656</td>
</tr>
<tr>
<td>XI</td>
<td><strong>UNDERSTANDING GEOMETRY</strong></td>
<td>657</td>
</tr>
</tbody>
</table>
## 38 Representing Images and Geometry

38.1 Introduction .......................................................... 659
38.2 Homogeneous and Heterogeneous Coordinates ................. 660
38.3 2D Image Transformations ....................................... 661
38.4 Lines and Planes in Homogeneous Coordinates ................ 668
38.5 Image Warping ......................................................... 670
38.6 Implicit Image Representations .................................. 671
38.7 Concluding Remarks ................................................ 673

## 39 Camera Modeling and Calibration

39.1 Introduction .......................................................... 675
39.2 3D Camera Projections in Homogeneous Coordinates ....... 676
39.3 Camera-Intrinsic Parameters ..................................... 678
39.4 Camera-Extrinsic Parameters .................................... 683
39.5 Full Camera Model ..................................................... 685
39.6 A Few Concrete Examples .......................................... 686
39.7 Camera Calibration .................................................... 692
39.8 Concluding Remarks ................................................ 699

## 40 Stereo Vision

40.1 Introduction .......................................................... 701
40.2 Stereo Cues ............................................................... 702
40.3 Model-Based Methods ............................................... 706
40.4 Learning-Based Methods .......................................... 717
40.5 Evaluation ............................................................... 719
40.6 Concluding Remarks ................................................ 719

## 41 Homographies

41.1 Introduction .......................................................... 721
41.2 Homography ............................................................. 722
41.3 Creating Image Panoramas ......................................... 727
41.4 Concluding Remarks ................................................ 730

## 42 Single View Metrology

42.1 Introduction .......................................................... 731
42.2 A Few Notes about Perception of Depth by Humans ......... 732
42.3 Linear Perspective ....................................................... 735
42.4 Measuring Heights Using Parallel Lines ......................... 741
42.5 3D Metrology from a Single View ............................... 749
42.6 Camera Calibration from Vanishing Points ...................... 753
42.7 Concluding Remarks ................................................ 755

## 43 Learning to Estimate Depth from a Single Image

.......................................................... 757
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>43.1</td>
<td>Introduction</td>
<td>757</td>
</tr>
<tr>
<td>43.2</td>
<td>Monocular Depth Cues</td>
<td>757</td>
</tr>
<tr>
<td>43.3</td>
<td>3D Representations</td>
<td>758</td>
</tr>
<tr>
<td>43.4</td>
<td>Supervised Methods for Depth from a Single Image</td>
<td>761</td>
</tr>
<tr>
<td>43.5</td>
<td>Unsupervised Methods for Depth from a Single Image</td>
<td>764</td>
</tr>
<tr>
<td>43.6</td>
<td>Concluding Remarks</td>
<td>767</td>
</tr>
<tr>
<td>44</td>
<td>Multiview Geometry and Structure from Motion</td>
<td>769</td>
</tr>
<tr>
<td>44.1</td>
<td>Introduction</td>
<td>769</td>
</tr>
<tr>
<td>44.2</td>
<td>Structure from Motion</td>
<td>769</td>
</tr>
<tr>
<td>44.3</td>
<td>Sparse SFM</td>
<td>771</td>
</tr>
<tr>
<td>44.4</td>
<td>Concluding Remarks</td>
<td>780</td>
</tr>
<tr>
<td>45</td>
<td>Radiance Fields</td>
<td>783</td>
</tr>
<tr>
<td>45.1</td>
<td>Introduction</td>
<td>783</td>
</tr>
<tr>
<td>45.2</td>
<td>What is a Radiance Field?</td>
<td>784</td>
</tr>
<tr>
<td>45.3</td>
<td>Representing Radiance Fields With Parameterized Functions</td>
<td>787</td>
</tr>
<tr>
<td>45.4</td>
<td>Rendering Radiance Fields</td>
<td>789</td>
</tr>
<tr>
<td>45.5</td>
<td>Fitting a Radiance Field to Explain a Scene</td>
<td>793</td>
</tr>
<tr>
<td>45.6</td>
<td>Beyond Radiance Fields: The Rendering Equation</td>
<td>797</td>
</tr>
<tr>
<td>45.7</td>
<td>Concluding Radiance Fields: The Rendering Equation</td>
<td>798</td>
</tr>
<tr>
<td>XII</td>
<td>UNDERSTANDING MOTION</td>
<td>799</td>
</tr>
<tr>
<td>46</td>
<td>Motion Estimation</td>
<td>801</td>
</tr>
<tr>
<td>46.1</td>
<td>Introduction</td>
<td>801</td>
</tr>
<tr>
<td>46.2</td>
<td>Motion Perception in the Human Visual System</td>
<td>802</td>
</tr>
<tr>
<td>46.3</td>
<td>Matching-Based Motion Estimation</td>
<td>804</td>
</tr>
<tr>
<td>46.4</td>
<td>Does the Human Visual System Use Matching to Estimate Motion?</td>
<td>808</td>
</tr>
<tr>
<td>46.5</td>
<td>Concluding Remarks</td>
<td>811</td>
</tr>
<tr>
<td>47</td>
<td>3D Motion and Its 2D Projection</td>
<td>813</td>
</tr>
<tr>
<td>47.1</td>
<td>Introduction</td>
<td>813</td>
</tr>
<tr>
<td>47.2</td>
<td>3D Motion and Its 2D Projection</td>
<td>813</td>
</tr>
<tr>
<td>47.3</td>
<td>Concluding Remarks</td>
<td>822</td>
</tr>
<tr>
<td>48</td>
<td>Optical Flow Estimation</td>
<td>823</td>
</tr>
<tr>
<td>48.1</td>
<td>Introduction</td>
<td>823</td>
</tr>
<tr>
<td>48.2</td>
<td>2D Motion Field and Optical Flow</td>
<td>823</td>
</tr>
<tr>
<td>48.3</td>
<td>Model-Based Approaches</td>
<td>826</td>
</tr>
<tr>
<td>48.4</td>
<td>Concluding Remarks</td>
<td>834</td>
</tr>
<tr>
<td>49</td>
<td>Learning to Estimate Motion</td>
<td>835</td>
</tr>
<tr>
<td>49.1</td>
<td>Introduction</td>
<td>835</td>
</tr>
</tbody>
</table>
## Contents

XIII UNDERSTANDING VISION WITH LANGUAGE ......................................................... 841

50 Object Recognition ................................................................. 843
  50.1 Introduction ................................................................. 843
  50.2 A Few Notes About Object Recognition in Humans ................. 844
  50.3 Image Classification ......................................................... 847
  50.4 Object Localization .......................................................... 854
  50.5 Class Segmentation ............................................................ 863
  50.6 Instance Segmentation ......................................................... 865
  50.7 Concluding Remarks .......................................................... 867

51 Vision and Language .......................................................... 869
  51.1 Introduction ................................................................. 869
  51.2 Background: Representing Text as Tokens .......................... 869
  51.3 Learning Visual Representations from Language Supervision .... 871
  51.4 Translating between Images and Text .................................... 877
  51.5 Text as a Visual Representation ............................................. 882
  51.6 Visual Question Answering .................................................. 883
  51.7 Concluding Remarks .......................................................... 884

XIV ON RESEARCH, WRITING AND SPEAKING ......................................................... 885

52 How to Do Research .......................................................... 887
  52.1 Introduction ................................................................. 887
  52.2 Research Advice .............................................................. 887
  52.3 Concluding Remarks .......................................................... 891

53 How to Write Papers .......................................................... 893
  53.1 Introduction ................................................................. 893
  53.2 Organization ................................................................. 894
  53.3 General Writing Tips ........................................................ 896
  53.4 Concluding Remarks .......................................................... 901

54 How to Give Talks ............................................................. 903
  54.1 Introduction ................................................................. 903
  54.2 Very Short Talks (2 – 10 minutes) ...................................... 903
  54.3 Preparation ................................................................. 904
  54.4 Nervousness ................................................................. 905
  54.5 Your Distracted Audience .................................................... 905
  54.6 Ways to Engage the Audience .............................................. 905
  54.7 Show Yourself to the Audience ............................................ 906