Contents

PREFACE

GLOSSARY OF NOTATIONS

1 VIDEO FORMATION, PERCEPTION, AND REPRESENTATION

1.1 Color Perception and Specification  2
  1.1.1 Light and Color, 2
  1.1.2 Human Perception of Color, 3
  1.1.3 The Trichromatic Theory of Color Mixture, 4
  1.1.4 Color Specification by Tristimulus Values, 5
  1.1.5 Color Specification by Luminance and Chrominance Attributes, 6

1.2 Video Capture and Display  7
  1.2.1 Principles of Color Video Imaging, 7
  1.2.2 Video Cameras, 8
  1.2.3 Video Display, 10
  1.2.4 Composite versus Component Video, 11
  1.2.5 Gamma Correction, 11

1.3 Analog Video Raster  12
  1.3.1 Progressive and Interlaced Scan, 12
  1.3.2 Characterization of a Video Raster, 14
# Contents

1.4 Analog Color Television Systems 16
   1.4.1 Spatial and Temporal Resolution, 16
   1.4.2 Color Coordinate, 17
   1.4.3 Signal Bandwidth, 19
   1.4.4 Multiplexing of Luminance, Chrominance, and Audio, 19
   1.4.5 Analog Video Recording, 21

1.5 Digital Video 22
   1.5.1 Notation, 22
   1.5.2 ITU-R BT.601 Digital Video, 23
   1.5.3 Other Digital Video Formats and Applications, 26
   1.5.4 Digital Video Recording, 28
   1.5.5 Video Quality Measure, 28

1.6 Summary 30

1.7 Problems 31

1.8 Bibliography 32

2 FOURIER ANALYSIS OF VIDEO SIGNALS AND
FREQUENCY RESPONSE OF THE HUMAN VISUAL SYSTEM 33

2.1 Multidimensional Continuous-Space Signals and Systems 33

2.2 Multidimensional Discrete-Space Signals and Systems 36

2.3 Frequency Domain Characterization of Video Signals 38
   2.3.1 Spatial and Temporal Frequencies, 38
   2.3.2 Temporal Frequencies Caused by Linear Motion, 40

2.4 Frequency Response of the Human Visual System 42
   2.4.1 Temporal Frequency Response and Flicker Perception, 43
   2.4.2 Spatial Frequency Response, 45
   2.4.3 Spatiotemporal Frequency Response, 46
   2.4.4 Smooth Pursuit Eye Movement, 48

2.5 Summary 50

2.6 Problems 51

2.7 Bibliography 52

3 VIDEO SAMPLING 53

3.1 Basics of the Lattice Theory 54

3.2 Sampling over Lattices 59
   3.2.1 Sampling Process and Sampled-Space Fourier Transform, 60
   3.2.2 The Generalized Nyquist Sampling Theorem, 61
   3.2.3 Sampling Efficiency, 63
Contents

3.2.4 Implementation of the Prefilter and Reconstruction Filter, 65
3.2.5 Relation between Fourier Transforms over Continuous, Discrete, and Sampled Spaces, 66

3.3 Sampling of Video Signals 67
3.3.1 Required Sampling Rates, 67
3.3.2 Sampling Video in Two Dimensions: Progressive versus Interlaced Scans, 69
3.3.3 Sampling a Raster Scan: BT.601 Format Revisited, 71
3.3.4 Sampling Video in Three Dimensions, 72
3.3.5 Spatial and Temporal Aliasing, 73

3.4 Filtering Operations in Cameras and Display Devices 76
3.4.1 Camera Apertures, 76
3.4.2 Display Apertures, 79

3.5 Summary 80

3.6 Problems 80

3.7 Bibliography 83

4 VIDEO SAMPLING RATE CONVERSION 84

4.1 Conversion of Signals Sampled on Different Lattices 84
4.1.1 Up-Conversion, 85
4.1.2 Down-Conversion, 87
4.1.3 Conversion between Arbitrary Lattices, 89
4.1.4 Filter Implementation and Design, and Other Interpolation Approaches, 91

4.2 Sampling Rate Conversion of Video Signals 92
4.2.1 Deinterlacing, 93
4.2.2 Conversion between PAL and NTSC Signals, 98
4.2.3 Motion-Adaptive Interpolation, 104

4.3 Summary 105

4.4 Problems 106

4.5 Bibliography 109

5 VIDEO MODELING 111

5.1 Camera Model 112
5.1.1 Pinhole Model, 112
5.1.2 CAHV Model, 114
5.1.3 Camera Motions, 116

5.2 Illumination Model 116
5.2.1 Diffuse and Specular Reflection, 116
Contents

5.2.2 Radiance Distribution under Differing Illumination and Reflection Conditions, 117
5.2.3 Changes in the Image Function Due to Object Motion, 119

5.3 Object Model  120
5.3.1 Shape Model, 121
5.3.2 Motion Model, 122

5.4 Scene Model  125

5.5 Two-Dimensional Motion Models  128
5.5.1 Definition and Notation, 128
5.5.2 Two-Dimensional Motion Models Corresponding to Typical Camera Motions, 130
5.5.3 Two-Dimensional Motion Corresponding to Three-Dimensional Rigid Motion, 133
5.5.4 Approximations of Projective Mapping, 136

5.6 Summary  137
5.7 Problems  138
5.8 Bibliography  139

6  TWO-DIMENSIONAL MOTION ESTIMATION  141

6.1 Optical Flow  142
6.1.1 Two-Dimensional Motion versus Optical Flow, 142
6.1.2 Optical Flow Equation and Ambiguity in Motion Estimation, 143

6.2 General Methodologies  145
6.2.1 Motion Representation, 146
6.2.2 Motion Estimation Criteria, 147
6.2.3 Optimization Methods, 151

6.3 Pixel-Based Motion Estimation  152
6.3.1 Regularization Using the Motion Smoothness Constraint, 153
6.3.2 Using a Multipoint Neighborhood, 153
6.3.3 Pel-Recursive Methods, 154

6.4 Block-Matching Algorithm  154
6.4.1 The Exhaustive Block-Matching Algorithm, 155
6.4.2 Fractional Accuracy Search, 157
6.4.3 Fast Algorithms, 159
6.4.4 Imposing Motion Smoothness Constraints, 161
6.4.5 Phase Correlation Method, 162
6.4.6 Binary Feature Matching, 163

6.5 Deformable Block-Matching Algorithms  165
6.5.1 Node-Based Motion Representation, 166
6.5.2 Motion Estimation Using the Node-Based Model, 167
Contents

6.6 Mesh-Based Motion Estimation 169
  6.6.1 Mesh-Based Motion Representation, 171
  6.6.2 Motion Estimation Using the Mesh-Based Model, 173

6.7 Global Motion Estimation 177
  6.7.1 Robust Estimators, 177
  6.7.2 Direct Estimation, 178
  6.7.3 Indirect Estimation, 178

6.8 Region-Based Motion Estimation 179
  6.8.1 Motion-Based Region Segmentation, 180
  6.8.2 Joint Region Segmentation and Motion Estimation, 181

6.9 Multiresolution Motion Estimation 182
  6.9.1 General Formulation, 182
  6.9.2 Hierarchical Block Matching Algorithm, 184

6.10 Application of Motion Estimation in Video Coding 187

6.11 Summary 188

6.12 Problems 189

6.13 Bibliography 191

7 THREE-DIMENSIONAL MOTION ESTIMATION 194

7.1 Feature-Based Motion Estimation 195
  7.1.1 Objects of Known Shape under Orthographic Projection, 195
  7.1.2 Objects of Known Shape under Perspective Projection, 196
  7.1.3 Planar Objects, 197
  7.1.4 Objects of Unknown Shape Using the Epipolar Line, 198

7.2 Direct Motion Estimation 203
  7.2.1 Image Signal Models and Motion, 204
  7.2.2 Objects of Known Shape, 206
  7.2.3 Planar Objects, 207
  7.2.4 Robust Estimation, 209

7.3 Iterative Motion Estimation 212

7.4 Summary 213

7.5 Problems 214

7.6 Bibliography 215

8 FOUNDATIONS OF VIDEO CODING 217

8.1 Overview of Coding Systems 218
  8.1.1 General Framework, 218
  8.1.2 Categorization of Video Coding Schemes, 219
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.2</td>
<td>Basic Notions in Probability and Information Theory</td>
<td>221</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Characterization of Stationary Sources</td>
<td>221</td>
</tr>
<tr>
<td>8.2.2</td>
<td>Entropy and Mutual Information for Discrete Sources</td>
<td>222</td>
</tr>
<tr>
<td>8.2.3</td>
<td>Entropy and Mutual Information for Continuous Sources</td>
<td>226</td>
</tr>
<tr>
<td>8.3</td>
<td>Information Theory for Source Coding</td>
<td>227</td>
</tr>
<tr>
<td>8.3.1</td>
<td>Bound for Lossless Coding</td>
<td>227</td>
</tr>
<tr>
<td>8.3.2</td>
<td>Bound for Lossy Coding</td>
<td>229</td>
</tr>
<tr>
<td>8.3.3</td>
<td>Rate-Distortion Bounds for Gaussian Sources</td>
<td>232</td>
</tr>
<tr>
<td>8.4</td>
<td>Binary Encoding</td>
<td>234</td>
</tr>
<tr>
<td>8.4.1</td>
<td>Huffman Coding</td>
<td>235</td>
</tr>
<tr>
<td>8.4.2</td>
<td>Arithmetic Coding</td>
<td>238</td>
</tr>
<tr>
<td>8.5</td>
<td>Scalar Quantization</td>
<td>241</td>
</tr>
<tr>
<td>8.5.1</td>
<td>Fundamentals</td>
<td>241</td>
</tr>
<tr>
<td>8.5.2</td>
<td>Uniform Quantization</td>
<td>243</td>
</tr>
<tr>
<td>8.5.3</td>
<td>Optimal Scalar Quantizer</td>
<td>244</td>
</tr>
<tr>
<td>8.6</td>
<td>Vector Quantization</td>
<td>248</td>
</tr>
<tr>
<td>8.6.1</td>
<td>Fundamentals</td>
<td>248</td>
</tr>
<tr>
<td>8.6.2</td>
<td>Lattice Vector Quantizer</td>
<td>251</td>
</tr>
<tr>
<td>8.6.3</td>
<td>Optimal Vector Quantizer</td>
<td>253</td>
</tr>
<tr>
<td>8.6.4</td>
<td>Entropy-Constrained Optimal Quantizer Design</td>
<td>255</td>
</tr>
<tr>
<td>8.7</td>
<td>Summary</td>
<td>257</td>
</tr>
<tr>
<td>8.8</td>
<td>Problems</td>
<td>259</td>
</tr>
<tr>
<td>8.9</td>
<td>Bibliography</td>
<td>261</td>
</tr>
</tbody>
</table>

9 WAVEFORM-BASED VIDEO CODING

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1</td>
<td>Block-Based Transform Coding</td>
<td>263</td>
</tr>
<tr>
<td>9.1.1</td>
<td>Overview</td>
<td>264</td>
</tr>
<tr>
<td>9.1.2</td>
<td>One-Dimensional Unitary Transform</td>
<td>266</td>
</tr>
<tr>
<td>9.1.3</td>
<td>Two-Dimensional Unitary Transform</td>
<td>269</td>
</tr>
<tr>
<td>9.1.4</td>
<td>The Discrete Cosine Transform</td>
<td>271</td>
</tr>
<tr>
<td>9.1.5</td>
<td>Bit Allocation and Transform Coding Gain</td>
<td>273</td>
</tr>
<tr>
<td>9.1.6</td>
<td>Optimal Transform Design and the KLT</td>
<td>279</td>
</tr>
<tr>
<td>9.1.7</td>
<td>DCT-Based Image Coders and the JPEG Standard</td>
<td>281</td>
</tr>
<tr>
<td>9.1.8</td>
<td>Vector Transform Coding</td>
<td>284</td>
</tr>
<tr>
<td>9.2</td>
<td>Predictive Coding</td>
<td>285</td>
</tr>
<tr>
<td>9.2.1</td>
<td>Overview</td>
<td>285</td>
</tr>
<tr>
<td>9.2.2</td>
<td>Optimal Predictor Design and Predictive Coding Gain</td>
<td>286</td>
</tr>
<tr>
<td>9.2.3</td>
<td>Spatial-Domain Linear Prediction</td>
<td>290</td>
</tr>
<tr>
<td>9.2.4</td>
<td>Motion-Compensated Temporal Prediction</td>
<td>291</td>
</tr>
</tbody>
</table>
9.3 Video Coding Using Temporal Prediction and Transform Coding 293
  9.3.1 Block-Based Hybrid Video Coding, 293
  9.3.2 Overlapped Block Motion Compensation, 296
  9.3.3 Coding Parameter Selection, 299
  9.3.4 Rate Control, 302
  9.3.5 Loop Filtering, 305

9.4 Summary 308

9.5 Problems 309

9.6 Bibliography 311

10 CONTENT-DEPENDENT VIDEO CODING 314

10.1 Two-Dimensional Shape Coding 314
  10.1.1 Bitmap Coding, 315
  10.1.2 Contour Coding, 318
  10.1.3 Evaluation Criteria for Shape Coding Efficiency, 323

10.2 Texture Coding for Arbitrarily Shaped Regions 324
  10.2.1 Texture Extrapolation, 324
  10.2.2 Direct Texture Coding, 325

10.3 Joint Shape and Texture Coding 326

10.4 Region-Based Video Coding 327

10.5 Object-Based Video Coding 328
  10.5.1 Source Model F2D, 330
  10.5.2 Source Models R3D and F3D, 332

10.6 Knowledge-Based Video Coding 336

10.7 Semantic Video Coding 338

10.8 Layered Coding System 339

10.9 Summary 342

10.10 Problems 343

10.11 Bibliography 344

11 SCALABLE VIDEO CODING 349

11.1 Basic Modes of Scalability 350
  11.1.1 Quality Scalability, 350
  11.1.2 Spatial Scalability, 353
  11.1.3 Temporal Scalability, 356
  11.1.4 Frequency Scalability, 356
Contents

11.1.5 Combination of Basic Schemes, 357
11.1.6 Fine-Granularity Scalability, 357

11.2 Object-Based Scalability 359

11.3 Wavelet-Transform-Based Coding 361
   11.3.1 Wavelet Coding of Still Images, 363
   11.3.2 Wavelet Coding of Video, 367

11.4 Summary 370

11.5 Problems 370

11.6 Bibliography 371

12 STEREO AND MULTIVIEW SEQUENCE PROCESSING 374

12.1 Depth Perception 375
   12.1.1 Binocular Cues—Stereopsis, 375
   12.1.2 Visual Sensitivity Thresholds for Depth Perception, 375

12.2 Stereo Imaging Principle 377
   12.2.1 Arbitrary Camera Configuration, 377
   12.2.2 Parallel Camera Configuration, 379
   12.2.3 Converging Camera Configuration, 381
   12.2.4 Epipolar Geometry, 383

12.3 Disparity Estimation 385
   12.3.1 Constraints on Disparity Distribution, 386
   12.3.2 Models for the Disparity Function, 387
   12.3.3 Block-Based Approach, 388
   12.3.4 Two-Dimensional Mesh-Based Approach, 388
   12.3.5 Intra-Line Edge Matching Using Dynamic Programming, 391
   12.3.6 Joint Structure and Motion Estimation, 392

12.4 Intermediate View Synthesis 393

12.5 Stereo Sequence Coding 396
   12.5.1 Block-Based Coding and MPEG-2 Multiview Profile, 396
   12.5.2 Incomplete Three-Dimensional Representation of Multiview Sequences, 398
   12.5.3 Mixed-Resolution Coding, 398
   12.5.4 Three-Dimensional Object-Based Coding, 399
   12.5.5 Three-Dimensional Model-Based Coding, 400

12.6 Summary 400

12.7 Problems 402

12.8 Bibliography 403
13 VIDEO COMPRESSION STANDARDS

13.1 Standardization 406
   13.1.1 Standards Organizations, 406
   13.1.2 Requirements for a Successful Standard, 409
   13.1.3 Standard Development Process, 411
   13.1.4 Applications for Modern Video Coding Standards, 412

13.2 Video Telephony with H.261 and H.263 413
   13.2.1 H.261 Overview, 413
   13.2.2 H.263 Highlights, 416
   13.2.3 Comparison, 420

13.3 Standards for Visual Communication Systems 421
   13.3.1 H.323 Multimedia Terminals, 421
   13.3.2 H.324 Multimedia Terminals, 422

13.4 Consumer Video Communications with MPEG-1 423
   13.4.1 Overview, 423
   13.4.2 MPEG-1 Video, 424

13.5 Digital TV with MPEG-2 426
   13.5.1 Systems, 426
   13.5.2 Audio, 426
   13.5.3 Video, 427
   13.5.4 Profiles, 435

13.6 Coding of Audiovisual Objects with MPEG-4 437
   13.6.1 Systems, 437
   13.6.2 Audio, 441
   13.6.3 Basic Video Coding, 442
   13.6.4 Object-Based Video Coding, 445
   13.6.5 Still Texture Coding, 447
   13.6.6 Mesh Animation, 447
   13.6.7 Face and Body Animation, 448
   13.6.8 Profiles, 451
   13.6.9 Evaluation of Subjective Video Quality, 454

13.7 Video Bit Stream Syntax 454

13.8 Multimedia Content Description Using MPEG-7 458
   13.8.1 Overview, 458
   13.8.2 Multimedia Description Schemes, 459
   13.8.3 Visual Descriptors and Description Schemes, 461

13.9 Summary 465

13.10 Problems 466

13.11 Bibliography 467
14 ERROR CONTROL IN VIDEO COMMUNICATIONS

14.1 Motivation and Overview of Approaches 473

14.2 Typical Video Applications and Communication Networks 476
  14.2.1 Categorization of Video Applications, 476
  14.2.2 Communication Networks, 479

14.3 Transport-Level Error Control 485
  14.3.1 Forward Error Correction, 485
  14.3.2 Error-Resilient Packetization and Multiplexing, 486
  14.3.3 Delay-Constrained Retransmission, 487
  14.3.4 Unequal Error Protection, 488

14.4 Error-Resilient Encoding 489
  14.4.1 Error Isolation, 489
  14.4.2 Robust Binary Encoding, 490
  14.4.3 Error-Resilient Prediction, 492
  14.4.4 Layered Coding with Unequal Error Protection, 493
  14.4.5 Multiple-Description Coding, 494
  14.4.6 Joint Source and Channel Coding, 498

14.5 Decoder Error Concealment 498
  14.5.1 Recovery of Texture Information, 500
  14.5.2 Recovery of Coding Modes and Motion Vectors, 501
  14.5.3 Syntax-Based Repair, 502

14.6 Encoder–Decoder Interactive Error Control 502
  14.6.1 Coding-Parameter Adaptation Based on Channel Conditions, 503
  14.6.2 Reference Picture Selection Based on Feedback Information, 503
  14.6.3 Error Tracking Based on Feedback Information, 504
  14.6.4 Retransmission without Waiting, 504

14.7 Error-Resilience Tools in H.263 and MPEG-4 505
  14.7.1 Error-Resilience Tools in H.263, 505
  14.7.2 Error-Resilience Tools in MPEG-4, 508

14.8 Summary 509

14.9 Problems 511

14.10 Bibliography 513

15 STREAMING VIDEO OVER THE INTERNET AND WIRELESS IP NETWORKS 519

15.1 Architecture for Video Streaming Systems 520

15.2 Video Compression 522
15.3 Application-Layer QoS Control for Streaming Video 522
   15.3.1 Congestion Control, 522
   15.3.2 Error Control, 525

15.4 Continuous Media Distribution Services 529
   15.4.1 Network Filtering, 529
   15.4.2 Application-Level Multicast, 531
   15.4.3 Content Replication, 532

15.5 Streaming Servers 533
   15.5.1 Real-Time Operating System, 534
   15.5.2 Storage System, 537

15.6 Media Synchronization 539

15.7 Protocols for Streaming Video 542
   15.7.1 Transport Protocols, 543
   15.7.2 Session Control Protocol: RTSP, 545

15.8 Streaming Video over Wireless IP Networks 546
   15.8.1 Network-Aware Applications, 548
   15.8.2 Adaptive Service, 549

15.9 Summary 554

15.10 Bibliography 555

APPENDIX A: DETERMINATION OF SPATIAL–TEMPORAL GRADIENTS 562
   A.1 First- and Second-Order Gradient 562
   A.2 Sobel Operator 563
   A.3 Difference of Gaussian Filters 563

APPENDIX B: GRADIENT DESCENT METHODS 565
   B.1 First-Order Gradient Descent Method 565
   B.2 Steepest Descent Method 566
   B.3 Newton’s Method 566
   B.4 Newton-Ralphson Method 567
   B.5 Bibliography 567

APPENDIX C: GLOSSARY OF ACRONYMS 568

APPENDIX D: ANSWERS TO SELECTED PROBLEMS 575