Contents

Preface **xvii**

Chapter 1 Introduction **1**
1.1 Image Engineering and Its Development 2
1.1.1 Basic Concepts and Entire Framework 2
1.1.2 Review of the Development of Image Technology 4
1.1.2.1 A Closed Survey Series of Image Technology 4
1.1.2.2 In Progress Survey Series of Image Engineering 4
1.2 Image Understanding and Related Disciplines 9
1.2.1 Image Understanding 9
1.2.2 Computer Vision 10
1.2.2.1 Research Methods 10
1.2.2.2 Realization of Engineering Methods 11
1.2.2.3 Research Objectives 11
1.2.2.4 The Relationship between Image Understanding and Computer Vision 12
1.2.3 Other Related Disciplines 13
1.2.3.1 Artificial Intelligence 13
1.2.3.2 Machine Learning and Deep Learning 13
1.2.3.3 Machine Vision/Robot Vision 14
1.2.3.4 Pattern Recognition 14
1.2.3.5 Computer Graphics 14
1.3 The Theoretical Framework of Image Understanding 15
1.3.1 Marr's Theory of Visual Computation 15
1.3.1.1 Vision is a Complex Information Processing Process 15
1.3.1.2 Three Key Elements of Visual Information Processing 16
1.3.1.3 Three-Level Internal Representation of Visual Information 18
1.3.1.4 Visual Information Understanding is Organized in the Form of Functional Modules 20
1.3.1.5 The Formal Representation of Computational Theory Must Take into Account Constraints 21
1.3.2 Improvements to Marr's Theoretical Framework 22
1.3.3 Discussion on Marr's Reconstruction Theory 24
1.3.3.1 Problems Related to Reconstruction Theory 24
1.3.3.2 Representation without Reconstruction 25
1.3.4 Research on the New Theoretical Framework 27
1.3.4.1 Knowledge-Based Theoretical Framework 27
1.3.4.2 Active Vision Theory Framework 28
1.4 Characteristics of This Book 29
1.4.1 Writing Motivation 29
1.4.2 Material Selection and Contents 31
1.4.3 Structure and Arrangement 32
References 33

Chapter 2 Camera Model and Calibration **35**
2.1 Linear Camera Model 36
2.1.1 Imaging Transformation 36
2.1.1.1 Various Coordinate Systems 36
2.1.1.2 Imaging Model 36
2.1.1.3 Perspective Transformation 37
2.1.1.4 Telecentric Imaging and Supercentric Imaging 38
2.1.1.5 Homogeneous Coordinates 40
2.1.1.6 Inverse Perspective Transformation 41
2.1.2 Approximate Projection Modes 42
2.1.2.1 Orthogonal Projection 42
2.1.2.2 Weak Perspective Projection 43
2.1.2.3 Parallel Perspective Projection 44
2.1.2.4 Comparison of Various Approximate Modes and Perspective Projection 45
2.1.3 A General Camera Model 45
2.2 Nonlinear Camera Model 47
2.2.1 Type of Distortion 48

2.2.1.1 Radial Distortion 48
2.2.1.2 Tangential Distortion 49
2.2.1.3 Eccentric Distortion 50
2.2.1.4 Thin Prism Distortion 50
2.2.2 A Complete Imaging Model 50
2.3 Camera Calibration 52
2.3.1 Basic Calibration Procedure 52
2.3.2 Camera Internal and External Parameters 53
2.3.2.1 External Parameters 54
2.3.2.2 Internal Parameters 54
2.3.2.3 Another Description of Internal and External Parameters 55
2.3.3 Nonlinear Camera Calibration 56
2.3.4 Classification of Calibration Methods 57
2.4 Traditional Calibration Methods 57
2.4.1 Basic Steps and Parameters 59
2.4.2 Two-Stage Calibration Method 59
2.4.3 Precision Improvement 62
2.5 Self-Calibration Methods 63
2.5.1 Basic Idea 63
2.5.2 A Practical Method 64
2.6 Some Recent Developments and Further Research 66
2.6.1 Calibration of Structured Light Active Vision System 66
2.6.1.1 Projector Model and Calibration 67
2.6.1.2 Pattern Separation 68
2.6.1.3 Calculation of Homography Matrix 69
2.6.1.4 Calculation of Calibration Parameters 71
2.6.2 U-Net Network for Medical Image Segmentation 73
2.6.2.1 Lane Line Detection and Data Screening 74
2.6.2.2 Optimizing Reprojection Error 75
References 76

Chapter 3 Stereo Vision 77
3.1 Depth Imaging and Depth Image 78
3.1.1 Depth Image and Grayscale Image 78
3.1.2 Intrinsic Image and Non-Intrinsic Image 79
3.1.3 Depth Imaging Modes 79
3.2 Binocular Imaging Modes 81
3.2.1 Binocular Horizontal Mode 81
3.2.1.1 Parallax and Depth 81
3.2.1.2 Angular Scanning Imaging 83
3.2.2 Binocular Convergence Horizontal Mode 84
3.2.2.1 Parallax and Depth 84
3.2.2.2 Image Rectification 85
3.2.3 Binocular Axial Mode 87
3.3 Binocular Stereo Matching Based on Region 88
3.3.1 Template Matching 88
3.3.1.1 Basic Method 89
3.3.1.2 Using Geometric Hashing 91
3.3.2 Stereo Matching 92
3.3.2.1 Epipolar Line Constraint 93
3.3.2.2 Essential Matrix and Fundamental Matrix 95
3.3.2.3 Calculation of Optical Properties 97
3.4 Binocular Stereo Matching Based on Feature 98
3.4.1 Basic Steps 98
3.4.1.1 Matching Using Edge Points 98
3.4.1.2 Matching Using Zero-Crossing Points 99
3.4.1.3 Feature Point Depth 100
3.4.1.4 Sparse Matching Points 101
3.4.2 Scale-Invariant Feature Transformation 101
3.4.3 Speed-Up Robust Feature 104
3.4.3.1 Determine the Point of Interest Based on the Hessian Matrix 104
3.4.3.2 Scale Space Representation 105

3.4.3.3 Description and Matching of Points of Interest 107
3.5 Some Recent Developments and Further Research 110
3.5.1 Biocular and Stereopsis 110
3.5.1.1 Biocular and Binocular 110
3.5.1.2 From Monocular to Binocular 111
3.5.2 Stereo Matching Methods Based on Deep Learning 112
3.5.2.1 Methods Using Image Pyramid Networks 112
3.5.2.2 Methods Using Siamese Networks 113
3.5.2.3 Methods Using Generative Adversarial Networks 113
3.5.3 Matching Based on Feature Cascade CNN 114
References 115

Chapter 4 Generalized Matching 117
4.1 Matching Overview 118
4.1.1 Matching Strategies and Categories 118
4.1.2 Matching and Registration 120
4.1.2.1 Registration Technology 120
4.1.2.2 Inertia Equivalent Ellipse Matching 121
4.1.3 Matching Evaluation 123
4.2 Object Matching 124
4.2.1 Matching metrics 124
4.2.1.1 Hausdorff Distance 125
4.2.1.2 Structural Matching Metrics 126
4.2.2 Corresponding Point Matching 127
4.2.3 String Matching 129
4.2.4 Shape Matrix Matching 129
4.3 Dynamic Pattern Matching 130
4.3.1 Matching Process 131
4.3.2 Absolute Pattern and Relative Pattern 131
4.4 Relationship Matching 134
4.4.1 Objects and Relational Representations 135
4.4.2 Connection Relationship Matching 137
4.4.3 Matching Process 138
4.5 Graph Isomorphism Matching 139
4.5.1 Introduction to Graph Theory 139
4.5.1.1 Basic Definitions 139
4.5.1.2 Geometric Representation of Graphs 140
4.5.1.3 Subgraphs 141
4.5.2 Graph Isomorphism and Matching 142
4.5.2.1 Identity and Isomorphism of Graphs 142
4.5.2.2 Judgment for Isomorphism 143
4.6 Some Recent Developments and Further Research 145
4.6.1 Image Registration and Matching 145
4.6.1.1 Heterogeneous Remote Sensing Image Registration Based on Feature Matching 145
4.6.1.2 Image Matching Based on Spatial Relation Reasoning 146
4.6.2 Multimodal Image Matching 148
4.6.2.1 Region-Based Techniques 148
4.6.2.2 Feature-Based Techniques 149
References 151

Chapter 5 Scene Analysis and Semantic Interpretation 155
5.1 Overview of Scene Understanding 155
5.1.1 Scenario Analysis 156
5.1.2 Scene Awareness Hierarchy 156
5.1.3 Scene Semantic Interpretation 158
5.2 Fuzzy Inference 158
5.2.1 Fuzzy Sets and Fuzzy Operations 159
5.2.2 Fuzzy Inference Methods 160
5.2.2.1 Basic Model 160
5.2.2.2 Fuzzy Combination 161
5.2.2.3 Defuzzification 163

5.3 Predicate Logical System 163
5.3.1 Predicate Calculus Rules 164
5.3.2 Inference by Theorem Proving 166
5.4 Scene Object Labeling 171
5.4.1 Labeling Methods and Key Elements 171
5.4.2 Discrete Relaxation Labeling 172
5.4.3 Probabilistic Relaxation Labels 173
5.5 Scene Classification 174
5.5.1 Models of Bag of Words/Bag of Features 175
5.5.2 pLSA Model 177
5.5.2.1 Model Description 178
5.5.2.2 Model Calculation 179
5.5.2.3 Model Application Example 180
5.5.3 LDA Model 182
5.5.3.1 Basic LDA Model 182
5.5.3.2 SLDA Model 184
5.6 Some Recent Developments and Further Research 185
5.6.1 Interpretation of Remote Sensing Images 185
5.6.1.1 Classification of Remote Sensing Image Interpretation Methods 185
5.6.1.2 Knowledge Graph for Remote Sensing Image Interpretation 187
5.6.2 Hybrid Enhanced Visual Cognition 188
5.6.2.1 From Computer Vision Perception to Computer Vision Cognition 189
5.6.2.2 Hybrid Enhanced Visual Cognition Related Technologies 190
References 192

Chapter 6 Multi-Sensor Image Information Fusion 195
6.1 Overview of Information Fusion 195
6.1.1 Multi-Sensor Information Fusion 196
6.1.2 Information Fusion Level 197
6.1.3 Active Vision and Active Fusion 198
6.2 Image Fusion 199
6.2.1 Main Steps of Image Fusion 199
6.2.1.1 Image Preprocessing 199
6.2.1.2 Image Registration 199
6.2.1.3 Image Information Fusion 200
6.2.2 Three Levels of Image Fusion 201
6.2.3 Evaluation of Image Fusion Effect 203
6.2.3.1 Subjective Evaluation 203
6.2.3.2 Objective Evaluation Based on Statistical Characteristics 204
6.2.3.3 Objective Evaluation Based on the Amount of Information 205
6.2.3.4 Evaluation According to the Purpose of Fusion 207
6.3 Pixel-Level Fusion Methods 207
6.3.1 Basic Fusion Methods 207
6.3.1.1 Weighted Average Fusion Method 208
6.3.1.2 Pyramid Fusion Method 208
6.3.1.3 Wavelet Transform Fusion Method 209
6.3.1.4 HSI Transform Fusion Method 209
6.3.1.5 PCA Transform Fusion Method 210
6.3.2 Combining Various Fusion Methods 210
6.3.2.1 Problems with a Single-Type Fusion Method 210
6.3.2.2 Fusion by Combining HSI Transform and Wavelet Transform 211
6.3.2.3 Fusion by Combining PCA Transform and Wavelet Transform 212
6.3.2.4 Performance of Combined Fusions 213
6.3.3 The Optimal Number of Decomposition Layers for Wavelet Fusion 214
6.3.4 Image Fusion Based on Compressed Sensing 216
6.3.5 Examples of Pixel-Level Fusion 218
6.3.5.1 Fusion of Different Exposure Images 218
6.3.5.2 Fusion of Different Focus Images 218
6.3.5.3 Fusion of Remote Sensing Images 219
6.3.5.4 Fusion of Visible Light Image and Infrared Image 219
6.3.5.5 Fusion of Visible Light Image and Millimeter Wave Radar Image 220
6.3.5.6 Fusion of CT Image and PET Image 220
6.3.5.7 Fusion of Dual Energy Transmission Image and Compton Backscatter Image 221

6.4 Feature-Level and Decision-Level Fusion Methods 222
6.4.1 Bayesian Method 222
6.4.2 Evidential Reasoning Method 224
6.5 Rough Set Theory in Decision-Level Fusion 226
6.5.1 Rough Set Definition 227
6.5.2 Rough Set Description 228
6.5.3 Fusion Based on Rough Sets 230
6.6 Some Recent Developments and Further Research 230
6.6.1 Spatial-Spectral Feature Extraction of Hyperspectral Images 230
6.6.1.1 Traditional Spatial-Spectral Feature Extraction Methods 231
6.6.1.2 Method for Extracting Spatial-Spectral Features Based on Deep Learning 232
6.6.2 Multi-Source Remote Sensing Image Fusion 233
6.6.2.1 Nine Multi-Source Remote Sensing Data Sources 233
6.6.2.2 Multi-Source Remote Sensing Image Fusion Literatures 236
6.6.2.3 Spatial-Spectral Fusion of Remote Sensing Images 237
6.6.2.4 Fusion with Deep Recurrent Residual Networks 237
References 240

Chapter 7 Content-Based Visual Information Retrieval 243
7.1 Principles of Image and Video Retrieval 243
7.1.1 Content-Based Retrieval 243
7.1.2 Achieving and Retrieval Flowchart 243
7.1.3 Multi Level Content Representation 246
7.2 Matching and Retrieval of Visual Features 248
7.2.1 Color Feature Matching 248
7.2.1.1 Histogram Intersection Method 249
7.2.1.2 Distance Method 249
7.2.1.3 Central Moment Method 249
7.2.1.4 Reference Color Table Method 249
7.2.2 Texture Feature Calculation 250
7.2.3 Multi-Scale Shape Features 252
7.2.4 Retrieval with Composite Features 253
7.2.4.1 Combination of Color and Texture Features 254
7.2.4.2 Combining Color, SIFT and CNN Features 256
7.3 Video Retrieval Based on Motion Features 256
7.3.1 Global Motion Features 257
7.3.2 Local Motion Features 258
7.4 Video Program Retrieval 259
7.4.1 News Video Structuring 260
7.4.1.1 Features of News Video 260
7.4.1.2 Main Speaker Close-up Shot Detection 260
7.4.1.3 Clustering of Main Speaker Close-up Shots 262
7.4.1.4 Announcer Shot Extraction 263
7.4.2 Video Ranking of Sports Games 264
7.4.2.1 Features of Sports Video 264
7.4.2.2 Structure of Table Tennis Competition Program 265
7.4.2.3 Object Detection and Tracking 266
7.4.2.4 Make the Brilliance Ranking 267
7.5 Semantic Classification Retrieval 268
7.5.1 Image Classification Based on Visual Keywords 269
7.5.1.1 Feature Selection 269
7.5.1.2 Image Classification 270
7.5.2 High-Level Semantics and Atmosphere 270
7.5.2.1 Five Atmospheric Semantics 271
7.5.2.2 Classification of Atmosphere 272
7.6 Some Recent Developments and Further Research 273
7.6.1 Deep Learning Based Cross-Modal Retrieval 273
7.6.2 Hashing in Image Retrieval 276
7.6.2.1 Supervised Hashing 276
7.6.2.2 Asymmetric Supervised Deep Discrete Hashing 277
7.6.2.3 Hashing in Cross Modal Image Retrieval 277
References 280

Chapter 8 Understanding Spatial-Temporal Behavior 283
8.1 Spatial-Temporal Technology 284
8.2 Spatial-Temporal Points of Interest 285
8.2.1 Detection of Spatial Points of Interest 286
8.2.2 Detection of Spatial-Temporal Points of Interest 287
8.3 Dynamic Trajectory Learning and Analysis 289
8.3.1 Overall Process 289
8.3.2 Automatic Scene Modeling 289
8.3.2.1 Object Tracking 290
8.3.2.2 Interest Point Detection 290
8.3.2.3 Activity Path Learning 291
8.3.3 Automated Activity Analysis 291
8.4 Action Classification and Recognition 293
8.4.1 Action Classification 294
8.4.1.1 Direct Classification 294
8.4.1.2 Time-State Model 294
8.4.1.3 Action Detection 294
8.4.2 Action Recognition 296
8.4.2.1 Holistic Recognition 296
8.4.2.2 Pose Modeling 296
8.4.2.3 Active Reconstruction 297
8.4.2.4 Interactive Activities 297
8.4.2.5 Group Activities 298
8.4.2.6 Scene Interpretation 298
8.5 Activity and Behavior Modeling 299
8.5.1 Action Modeling 300
8.5.1.1 Non-Parametric Modeling Methods 300
8.5.1.2 3-D Modeling Methods 301
8.5.1.3 Parametric Time Series Modeling Methods 303
8.5.2 Activity Modeling and Recognition 305
8.5.2.1 Graph Model 305
8.5.2.2 Synthesis Methods 307
8.5.2.3 Knowledge- and Logic-Based Methods 309
8.6 Joint Modeling of Actor and Action 310
8.6.1 Single-Label Actor-Action Recognition 311
8.6.2 Multi-Label Actor-Action Recognition 312
8.6.3 Actor-Action Semantic Segmentation 313
8.7 Some Recent Developments and Further Research 316
8.7.1 Behavior Recognition Using Joints 316
8.7.1.1 *Using CNN as Backbone* 317
8.7.1.2 *Using RNN as Backbone* 317
8.7.1.3 *Using GCN as Backbone* 318
8.7.1.4 *Using Hybrid Network as Backbone* 318
8.7.2 Detection of Video Anomalous Events 318
8.7.2.1 *Detection with Convolutional Auto-Encoder Block Learning* 319
8.7.2.2 *Detection Using One-Class Neural Network* 321
References 322

Index 325