Handbook of Image Processing and Computer Vision

Arcangelo Distante • Cosimo Distante

Handbook of Image Processing and Computer Vision

Volume 3: From Pattern to Object

Springer

Arcangelo Distante
Institute of Applied Sciences
and Intelligent Systems
Consiglio Nazionale delle Ricerche
Lecce, Italy

Cosimo Distante
Institute of Applied Sciences
and Intelligent Systems
Consiglio Nazionale delle Ricerche
Lecce, Italy

ISBN 978-3-030-42377-3 ISBN 978-3-030-42378-0 (eBook)
https://doi.org/10.1007/978-3-030-42378-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To my parents and my family, Maria and Maria Grazia—Arcangelo Distante

To my parents, to my wife Giovanna, and to my children Francesca and Davide—Cosimo Distante

Preface

In the last 20 years, several interdisciplinary researches in the fields of *physics, information technology and cybernetics, the numerical processing of Signals and Images, and electrical and electronic technologies* have led to the development of Intelligent Systems.

The so-called *Intelligent Systems (or Intelligent Agents)* represent the still more advanced and innovative frontier of research in the electronic and computer field, able to directly influence the quality of life, competitiveness, and production methods of companies, to monitor and evaluate the environmental impact, to make public service and management activities more efficient, and to protect people's safety.

The study of an intelligent system, regardless of the area of use, can be simplified into three essential components:

1. The *first* interacts with the environment for the acquisition of data of the domain of interest, using appropriate sensors (for the acquisition of Signals and Images);
2. The *second* analyzes and interprets the data collected by the first component, also using learning techniques to build/update adequate representations of the complex reality in which the system operates (Computational Vision);
3. The *third* chooses the most appropriate actions to achieve the objectives assigned to the intelligent system (choice of Optimal Decision Models) interacting with the first two components, and with human operators, in case of application solutions based on man–machine cooperative paradigms (the current evolution of automation including industrial one).

In this scenario of knowledge advancement for the development of Intelligent Systems, the information content of this manuscript is framed in which are reported the experiences of multi-year research and teaching of the authors, and of the scientific insights existing in the literature. In particular, the manuscript divided into three parts (volumes), deals with aspects of the sensory subsystem in order to perceive the environment in which an intelligent system is immersed and able to act autonomously.

The *first volume* describes the set of fundamental processes of artificial vision that lead to the formation of the digital image from energy. The phenomena of light propagation (Chaps. 1 and 2), the theory of color perception (Chap. 3), the impact of the optical system (Chap. 4), the aspects of transduction from luminous energy are analyzed (the optical flow) with electrical signal (of the photoreceptors), and aspects of electrical signal transduction (with continuous values) in discrete values (pixels), i.e., the conversion of the signal from analog to digital (Chap. 5). These first 5 chapters summarize the process of acquisition of the 3D scene, in symbolic form, represented numerically by the pixels of the digital image (2D projection of the 3D scene).

Chapter 6 describes the geometric, topological, quality, and perceptual information of the digital image. The metrics are defined, the aggregation and correlation modalities between pixels, useful for defining symbolic structures of the scene of higher level with respect to the pixel. The organization of the data for the different processing levels is described in Chap. 7 while in Chapter 8, the representation and description of the homogeneous structures of the scene is shown.

With Chapter 9 starts the description of the image processing algorithms, for the improvement of the visual qualities of the image, based on point, local, and global operators. Algorithms operating in the spatial domain and in the frequency domain are shown, highlighting with examples the significant differences between the various algorithms also from the point of view of the computational load.

The *second volume* begins with the chapter describing the boundary extraction algorithms based on local operators in the spatial domain and on filtering techniques in the frequency domain.

In Chap. 2 are presented the fundamental linear transformations that have immediate application in the field of image processing, in particular, to extract the essential characteristics contained in the images. These characteristics, which effectively summarize the global informational character of the image, are then used for the other image processing processes: classification, compression, description, etc. Linear transforms are also used, as global operators, to improve the visual qualities of the image (*enhancement*), to attenuate noise (*restoration*), or to reduce the dimensionality of the data (*data reduction*).

In Chap. 3, the geometric transformations of the images are described, necessary in different applications of the artificial vision, both to correct any geometric distortions introduced during the acquisition (for example, images acquired while the objects or the sensors are moving, as in the case of satellite and/or aerial acquisitions), or to introduce desired visual geometric effects. In both cases, the geometrical operator must be able to reproduce as accurately as possible the image with the same initial information content through the image resampling process.

In Chap. 4*Reconstruction of the degraded image (image restoration)*, a set of techniques are described that perform quantitative corrections on the image to compensate for the degradations introduced during the acquisition and transmission process. These degradations are represented by the fog or blurring effect caused by the optical system and by the motion of the object or the observer, by the noise caused by the opto-electronic system and by the nonlinear response of the sensors, by random noise due to atmospheric turbulence or, more generally, from the process of digitization and transmission. While the *enhancement* techniques tend to reduce the degradations present in the image in qualitative terms, improving their visual quality even when there is no knowledge of the degradation model, the *restoration* techniques are used instead to eliminate or quantitatively attenuate the degradations present in the image, starting also from the hypothesis of knowledge of degradation models.

Chapter 5, *Image Segmentation*, describes different segmentation algorithms, which is the process of dividing the image into homogeneous regions, where all the pixels that correspond to an object in the scene are grouped together. The grouping of pixels in regions is based on a homogeneity criterion that distinguishes them from one another. Segmentation algorithms based on criteria of similarity of pixel attributes (color, texture, etc.) or based on geometric criteria of spatial proximity of pixels (Euclidean distance, etc.) are reported. These criteria are not always valid, and in different applications, it is necessary to integrate other information in relation to the a priori knowledge of the application context (application domain). In this last case, the grouping of the pixels is based on comparing the hypothesized regions with the a priori modeled regions.

Chapter 6, *Detectors and descriptors of points of interest*, describes the most used algorithms to automatically detect significant structures (known as points of interest, corners, features) present in the image corresponding to stable physical parts of the scene. The ability of such algorithms is to detect and identify physical parts of the same scene in a repeatable way, even when the images are acquired under conditions of lighting variability and change of the observation point with possible change of the scale factor.

The third volume describes the artificial vision algorithms that detect objects in the scene, attempt their identification, 3D reconstruction, their arrangement and location with respect to the observer, and their eventual movement.

Chapter 1, *Object recognition*, describes the fundamental algorithms of artificial vision to automatically recognize the objects of the scene, essential characteristics of all systems of vision of living organisms. While a human observer also recognizes complex objects, apparently in an easy and timely manner, for a vision machine, the recognition process is difficult, requires considerable calculation time, and the results are not always optimal. Fundamental to the process of object recognition become the algorithms for selecting and extracting features. In various applications, it is possible to have an a priori knowledge of all the objects to be classified because we know the sample patterns (meaningful features) from which we can extract useful information for the decision to associate (*decision-making*) each individual of the population to a certain class. These sample patterns (*training set*) are used by the recognition system to learn significant information about the objects population (extraction of statistical parameters, relevant characteristics, etc.). The recognition process compares the features of the unknown objects to the model pattern features, in order to uniquely identify their class of membership. Over the years, there have been various disciplinary sectors (machine learning, image analysis, object recognition, information research, bioinformatics, biomedicine, intelligent data analysis, data mining, …) and the application sectors (robotics, remote sensing, artificial vision, …) for which different researchers have proposed different methods of recognition and developed different algorithms based on different classification models. Although the proposed algorithms have a unique purpose, they differ in the property attributed to the classes of objects (the clusters) and the model with which these classes are defined (connectivity, statistical distribution, density, …). The diversity of disciplines, especially between automatic data extraction (*data mining*) and machine learning (*machine learning*), has led to subtle differences, especially in the use of results and in terminology, sometimes contradictory, perhaps caused by the different objectives. For example, in data mining the dominant interest is automatic grouping extraction, in automatic classification the discriminating power of the pattern classes is fundamental. The topics of this chapter overlap between aspects related to machine learning and those of recognition based on statistical methods. For simplicity, the algorithms described are broken down according to the methods of classifying objects in *supervised methods* (based on deterministic, statistical, neural, and nonmetric models such as syntactic models and decision trees) and *non-supervised methods*, i.e., methods that do not use any prior knowledge to extract the classes to which the patterns belong.

In Chap. 2 *RBF, SOM, Hopfield and deep neural networks*, four different types of neural networks are described: *Radial Basis Functions*—RBF, *Self-Organizing Maps*—SOM, the Hopfield, and the deep neural networks. RBF uses a different approach in the design of a neural network based on the *hidden* layer (unique in the network) composed of neurons in which radial-based functions are defined, hence the name of *Radial Basis Functions*, and which performs a nonlinear transformation of the input data supplied to the network. These neurons are the basis for input data (vectors). The reason why a nonlinear transformation is used in the hidden layer, followed by a linear one in the output one, allows a pattern classification problem to operate in a much larger space (in nonlinear transformation from the input in the hidden one) and is more likely to be linearly separable than a small-sized space. From this observation, derives the reason why the hidden layer is generally larger than the input one (i.e., the number of hidden neurons is greater than the cardinality of the input signal).

The SOM network, on the other hand, has an unsupervised learning model and has the originality of autonomously grouping input data on the basis of their similarity without evaluating the convergence error with external information on the data. It is useful when there is no exact knowledge on the data to classify them. It is inspired by the topology of the brain cortex model considering the connectivity of the neurons and in particular, the behavior of an activated neuron and the influence with neighboring neurons that reinforce the connections compared to those further away that are becoming weaker.

With the Hopfield network, the learning model is supervised and with the ability to store information and retrieve it through even partial content of the original information. It presents its originality based on physical foundations that have revitalized the entire field of neural networks. The network is associated with an energy function to be minimized during its evolution with a succession of states, until reaching a final state corresponding to the minimum of the energy function. This feature allows it to be used to solve and set up an optimization problem in terms of the objective function to be associated with an energy function. The chapter concludes with the description of the *convolutional neural networks* (CNN), by now the most widespread since 2012, based on the deep learning architecture (*deep learning*).

In Chap. 3 *Texture Analysis*, the algorithms that characterize the texture present in the images are shown. Texture is an important component for the recognition of objects. In the field of image processing has been consolidated with the term texture, any geometric and repetitive arrangement of the levels of gray (or color) of an image. In this context, texture becomes an additional strategic component to solve the problem of object recognition, the segmentation of images, and the problems of synthesis. Some of the algorithms described are based on the mechanisms of human visual perception of texture. They are useful for the development of systems for the automatic analysis of the information content of an image obtaining a partitioning of the image in regions with different textures.

In Chap. 4 *3D Vision Paradigms* are reported the algorithms that analyze 2D images to reconstruct a scene typically of 3D objects. A 3D vision system that has the fundamental problem typical of inverse problems, i.e., from single 2D images, which are only a two-dimensional projection of the 3D world (partial acquisition), must be able to reconstruct the 3D structure of the observed scene and eventually define a relationship between the objects. 3D reconstruction takes place starting from 2D images that contain only partial information of the 3D world (loss of information from the projection 3D→2D) and possibly using the geometric and radiometric calibration parameters of the acquisition system. The mechanisms of human vision are illustrated, based also on the a priori prediction and knowledge of the world. In the field of artificial vision, the current trend is to develop 3D systems oriented to specific domains but with characteristics that go in the direction of imitating certain functions of the human visual system. 3D reconstruction methods are described that use multiple cameras observing the scene from multiple points of view, or sequences of time-varying images acquired from a single camera. Theories of vision are described, from the Gestalt laws to the paradigm of Marr's vision and the computational models of stereovision.

In Chap. 5 *Shape from Shading—(SfS)* are reported the algorithms to reconstruct the *shape* of the visible 3D surface using only the brightness variation information (*shading*, that is, the level variations of gray or colored) present in the image. The inverse problem of reconstructing the shape of the surface visible from the changes in brightness in the image is known as the *Shape from Shading* problem. The reconstruction of the visible surface should not be strictly understood as a 3D reconstruction of the surface. In fact, from a single point of the observation of the scene, a monocular vision system cannot estimate a distance measure between observer and visible object, so with the *SfS* algorithms, there is a nonmetric but qualitative reconstruction of the 3D surface. It is described the theory of the *SfS* based on the knowledge of the light source (direction and distribution), the model of reflectance of the scene, the observation point, and the geometry of the visible surface, which together contribute to the image formation process. The relationships between the light intensity values of the image and the geometry of the visible surface are derived (in terms of the orientation of the surface, point by point) under some lighting conditions and the reflectance model. Other 3D surface reconstruction algorithms based on the *Shape from xxx* paradigm are also described, where *xxx* can be *texture, structured light projected onto the surface to be reconstructed, or 2D images of the focused or defocused surface*.

In Chap. 6 *Motion Analysis*, the algorithms of perception of the dynamics of the scene are reported, analogous to what happens in the vision systems of different living beings. With motion analysis algorithms, it is possible to derive the 3D motion, almost in real time, from the analysis of sequences of time-varying 2D images.

Paradigms on movement analysis have shown that the perception of movement derives from the information of the objects evaluating the presence of occlusions, texture, contours, etc. The algorithms for the perception of the movement occurring in the physical reality and not the apparent movement are described. Different methods of movement analysis are analyzed from those with limited computational load such as those based on time-variant image difference to the more complex ones based on optical flow considering application contexts with different levels of motion entities and scene-environment with different complexities.

In the context of rigid bodies, from the motion analysis, derived from a sequence of time-variant images, are described the algorithms that, in addition to the movement (translation and rotation), estimate the reconstruction of the 3D structure of the scene and the distance of this structure by the observer. Useful information are obtained in the case of mobile observer (robot or vehicle) to estimate the collision time. In fact, the methods for solving the problem of 3D reconstruction of the scene are acquired by acquiring a sequence of images with a single camera whose intrinsic parameters remain constant even if not known (camera not calibrated) together with the non-knowledge of motion. The proposed methods are part of the problem of solving an inverse problem. Algorithms are described to reconstruct the 3D structure of the scene (and the motion), i.e., to calculate the coordinates of 3D points of the scene whose 2D projection is known in each image of the time-variant sequence.

Finally, in Chap. 7 *Camera Calibration and 3D Reconstruction*, the algorithms for calibrating the image acquisition system (normally a single camera and stereovision) are fundamental for detecting metric information (detecting an object's size or determining accurate measurements of object–observer distance) of the scene from the image. The various camera calibration methods are described that determine the relative *intrinsic parameters* (focal length, horizontal and vertical dimension of the single photoreceptor of the sensor, or the aspect ratio, the size of the matrix of the sensor, the coefficients of the radial distortion model, the coordinates of the main point or the optical center) and the *extrinsic parameters* that define the geometric transformation to pass from the reference system of the world to that of camera. The *epipolar geometry* introduced in Chap. 5 is described in this chapter to solve the problem of correspondence of homologous points in a stereo vision system with the two cameras calibrated and not. With the epipolar geometry is simplified the search for the homologous points between the stereo images introducing the *Essential matrix* and the *Fundamental matrix*. The algorithms for estimating these matrices are also described, known a priori the corresponding points of a calibration platform.

With epipolar geometry, the problem of searching for homologous points is reduced to mapping a point of an image on the corresponding epipolar line in the other image. It is possible to simplify the problem of correspondence through a one-dimensional point-to-point search between the stereo images. This is accomplished with the image alignment procedure, known as stereo image rectification. The different algorithms have been described; some based on the constraints of the epipolar geometry (non-calibrated cameras where the fundamental matrix includes the intrinsic parameters) and on the knowledge or not of the intrinsic and extrinsic parameters of calibrated cameras. Chapter 7 ends with the section of the 3D reconstruction of the scene in relation to the knowledge available to the stereo acquisition system. The triangulation procedures for the 3D reconstruction of the geometry of the scene without ambiguity are described, given the 2D projections of the homologous points of the stereo images, known the calibration parameters of the stereo system. If only the intrinsic parameters are known, the 3D geometry of the scene is reconstructed by estimating the extrinsic parameters of the system at less than a non-determinable scale factor. If the calibration parameters of the stereo system are not available but only the correspondences between the stereo images are known, the structure of the scene is recovered through an unknown homography transformation.

Francavilla Fontana, Italy — Arcangelo Distante
December 2020 — Cosimo Distante

Acknowledgments

We thank all the fellow researchers of the Department of Physics of Bari, of the Institute of Intelligent Systems for Automation of the CNR (National Research Council) of Bari, and of the Institute of Applied Sciences and Intelligent Systems “Eduardo Caianiello” of the Unit of Lecce, who have indicated errors and parts to be reviewed. We mention them in chronological order: Grazia Cicirelli, Marco Leo, Giorgio Maggi, Rosalia Maglietta, Annalisa Milella, Pierluigi Mazzeo, Paolo Spagnolo, Ettore Stella, and Nicola Veneziani. A thank you is addressed to Arturo Argentieri for the support on the graphic aspects of the figures and the cover. Finally, special thanks are given to Maria Grazia Distante who helped us realize the electronic composition of the volumes by verifying the accuracy of the text and the formulas.

Contents

1 Object Recognition 1
1.1 Introduction 1
1.2 Classification Methods 3
1.3 Prior Knowledge and Features Selection 4
1.4 Extraction of Significant Features 6
1.4.1 Feature Space 8
1.4.2 Selection of Significant Features 10
1.5 Interactive Method 17
1.6 Deterministic Method 17
1.6.1 Linear Discriminant Functions 19
1.6.2 Generalized Discriminant Functions 20
1.6.3 Fisher's Linear Discriminant Function 21
1.6.4 Classifier Based on Minimum Distance 28
1.6.5 Nearest-Neighbor Classifier 30
1.6.6 K-means Classifier 30
1.6.7 ISODATA Classifier 34
1.6.8 Fuzzy C-means Classifier 35
1.7 Statistical Method 37
1.7.1 MAP Classifier 37
1.7.2 Maximum Likelihood Classifier—ML 39
1.7.3 Other Decision Criteria 48
1.7.4 Parametric Bayes Classifier 48
1.7.5 Maximum Likelihood Estimation—MLE 49
1.7.6 Estimation of the Distribution Parameters with the Bayes Theorem 52
1.7.7 Comparison Between Bayesian Learning and Maximum Likelihood Estimation 56
1.8 Bayesian Discriminant Functions 57
1.8.1 Classifier Based on Gaussian Probability Density 58
1.8.2 Discriminant Functions for the Gaussian Density 61

1.9 Mixtures of Gaussian—MoG . . . 66
1.9.1 Parameters Estimation of the Gaussians Mixture with the Maximum Likelihood—ML . . . 68
1.9.2 Parameters Estimation of the Gaussians Mixture with Expectation–Maximization—EM . . . 69
1.9.3 EM Theory . . . 73
1.9.4 Nonparametric Classifiers . . . 76
1.10 Method Based on Neural Networks . . . 87
1.10.1 Biological Motivation . . . 87
1.10.2 Mathematical Model of the Neural Network . . . 89
1.10.3 Perceptron for Classification . . . 93
1.10.4 Linear Discriminant Functions and Learning . . . 100
1.11 Neural Networks . . . 109
1.11.1 Multilayer Perceptron—MLP . . . 109
1.11.2 Multilayer Neural Network for Classification . . . 111
1.11.3 Backpropagation Algorithm . . . 113
1.11.4 Learning Mode with Backpropagation . . . 118
1.11.5 Generalization of the MLP Network . . . 120
1.11.6 Heuristics to Improve Backpropagation . . . 121
1.12 Nonmetric Recognition Methods . . . 125
1.13 Decision Tree . . . 125
1.13.1 Algorithms for the Construction of Decision Trees . . . 127
1.14 ID3 Algorithm . . . 129
1.14.1 Entropy as a Measure of Homogeneity of the Samples . . . 129
1.14.2 Information Gain . . . 131
1.14.3 Other Partitioning Criteria . . . 135
1.15 C4.5 Algorithm . . . 137
1.15.1 Pruning Decision Tree . . . 138
1.15.2 Post-pruning Algorithms . . . 140
1.16 CART Algorithm . . . 143
1.16.1 Gini Index . . . 143
1.16.2 Advantages and Disadvantages of Decision Trees . . . 146
1.17 Hierarchical Methods . . . 148
1.17.1 Agglomerative Hierarchical Clustering . . . 149
1.17.2 Divisive Hierarchical Clustering . . . 151
1.17.3 Example of Hierarchical Agglomerative Clustering . . . 152
1.18 Syntactic Pattern Recognition Methods . . . 154
1.18.1 Formal Grammar . . . 156
1.18.2 Language Generation . . . 158
1.18.3 Types of Grammars . . . 161
1.18.4 Grammars for Pattern Recognition . . . 163
1.18.5 Notes on Other Methods of Syntactic Analysis . . . 171

1.19 String Recognition Methods . . . 172
1.19.1 String Matching . . . 172
1.19.2 Boyer–Moore String-Matching Algorithm . . . 174
1.19.3 Edit Distance . . . 183
1.19.4 String Matching with Error . . . 188
1.19.5 String Matching with Special Symbol . . . 190
References . . . 191

2 RBF, SOM, Hopfield, and Deep Neural Networks . . . 193
2.1 Introduction . . . 193
2.2 Cover Theorem on Pattern Separability . . . 194
2.3 The Problem of Interpolation . . . 196
2.4 Micchelli's Theorem . . . 199
2.5 Learning and Ill-Posed Problems . . . 200
2.6 Regularization Theory . . . 201
2.7 RBF Network Architecture . . . 205
2.8 RBF Network Solution . . . 207
2.9 Learning Strategies . . . 208
2.9.1 Centers Set and Randomly Selected . . . 208
2.9.2 Selection of Centers Using Clustering Techniques . . . 210
2.10 Kohonen Neural Network . . . 210
2.10.1 Architecture of the SOM Network . . . 211
2.10.2 SOM Applications . . . 219
2.11 Network Learning Vector Quantization-LVQ . . . 220
2.11.1 LVQ2 and LVQ3 Networks . . . 223
2.12 Recurrent Neural Networks . . . 223
2.12.1 Hopfield Network . . . 225
2.12.2 Application of Hopfield Network to Discrete States . . . 228
2.12.3 Continuous State Hopfield Networks . . . 233
2.12.4 Boltzmann Machine . . . 236
2.13 Deep Learning . . . 238
2.13.1 Deep Traditional Neural Network . . . 239
2.13.2 Convolutional Neural Networks-CNN . . . 240
2.13.3 Operation of a CNN Network . . . 248
2.13.4 Main Architectures of CNN Networks . . . 256
References . . . 257

3 Texture Analysis . . . 261
3.1 Introduction . . . 261
3.2 The Visual Perception of the Texture . . . 261
3.2.1 Julesz's Conjecture . . . 263
3.2.2 Texton Statistics . . . 264
3.2.3 Spectral Models of the Texture . . . 265

3.3 Texture Analysis and its Applications 265
3.4 Statistical Texture Methods 267
3.4.1 First-Order Statistics 267
3.4.2 Second-Order Statistics 268
3.4.3 Higher Order Statistics 268
3.4.4 Second-Order Statistics with Co-Occurrence Matrix 270
3.4.5 Texture Parameters Based on the Co-Occurrence Matrix 272
3.5 Texture Features Based on Autocorrelation 276
3.6 Texture Spectral Method 278
3.7 Texture Based on the Edge Metric 281
3.8 Texture Based on the Run Length Primitives 282
3.9 Texture Based on MRF, SAR, and Fractals Models 286
3.10 Texture by Spatial Filtering 290
3.10.1 Spatial Filtering with Gabor Filters 295
3.11 Syntactic Methods for Texture 302
3.12 Method for Describing Oriented Textures 303
3.12.1 Estimation of the Dominant Local Orientation 304
3.12.2 Texture Coherence 306
3.12.3 Intrinsic Images with Oriented Texture 307
3.13 Tamura's Texture Features 308
References 313

4 Paradigms for 3D Vision 315
4.1 Introduction to 3D Vision 315
4.2 Toward an Optimal 3D Vision Strategy 316
4.3 Toward the Marr's Paradigm 318
4.4 The Fundamentals of Marr's Theory 319
4.4.1 Primal Sketch 320
4.4.2 Toward a Perceptive Organization 325
4.4.3 The Gestalt Theory 326
4.4.4 From the Gestalt Laws to the Marr Theory 338
4.4.5 2.5D Sketch of Marr's Theory 342
4.5 Toward 3D Reconstruction of Objects 344
4.5.1 From Image to Surface: 2.5D Sketch Map Extraction 347
4.6 Stereo Vision 350
4.6.1 Binocular Fusion 351
4.6.2 Stereoscopic Vision 354
4.6.3 Stereopsis 356
4.6.4 Neurophysiological Evidence of Stereopsis 358
4.6.5 Depth Map from Binocular Vision 374
4.6.6 Computational Model for Binocular Vision 377

4.6.7 Simple Artificial Binocular System . . . 383
4.6.8 General Binocular System . . . 389
4.7 Stereo Vision Algorithms . . . 393
4.7.1 Point-Like Elementary Structures . . . 394
4.7.2 Local Elementary Structures and Correspondence Calculation Methods . . . 396
4.7.3 Sparse Elementary Structures . . . 406
4.7.4 PMF Stereo Vision Algorithm . . . 406
References . . . 410

5 Shape from Shading . . . 413
5.1 Introduction . . . 413
5.2 The Reflectance Map . . . 414
5.2.1 Gradient Space . . . 417
5.3 The Fundamental Relationship of Shape from Shading for Diffuse Reflectance . . . 420
5.4 Shape from Shading-SfS Algorithms . . . 423
5.4.1 Shape from Stereo Photometry with Calibration . . . 426
5.4.2 Uncalibrated Stereo Photometry . . . 433
5.4.3 Stereo Photometry with Calibration Sphere . . . 436
5.4.4 Limitations of Stereo Photometry . . . 439
5.4.5 Surface Reconstruction from the Orientation Map . . . 440
5.5 Shape from Texture . . . 447
5.6 Shape from Structured Light . . . 451
5.6.1 Shape from Structured Light with Binary Coding . . . 454
5.6.2 Gray Code Structured Lighting . . . 456
5.6.3 Pattern with Gray Level . . . 458
5.6.4 Pattern with Phase Modulation . . . 458
5.6.5 Pattern with Phase Modulation and Binary Code . . . 462
5.6.6 Methods Based on Colored Patterns . . . 462
5.6.7 Calibration of the Camera-Projector Scanning System . . . 463
5.7 Shape from (de)Focus . . . 465
5.7.1 Shape from Focus (SfF) . . . 465
5.7.2 Shape from Defocus (SfD) . . . 472
References . . . 477

6 Motion Analysis . . . 479
6.1 Introduction . . . 479
6.2 Analogy Between Motion Perception and Depth Evaluated with Stereo Vision . . . 482
6.3 Toward Motion Estimation . . . 484
6.3.1 Discretization of Motion . . . 487
6.3.2 Motion Estimation—Continuous Approach . . . 493

6.3.3 Motion Estimation—Discrete Approach 495
6.3.4 Motion Analysis from Image Difference 496
6.3.5 Motion Analysis from the Cumulative Difference of Images 496
6.3.6 Ambiguity in Motion Analysis 498
6.4 Optical Flow Estimation 499
6.4.1 Horn–Schunck Method 504
6.4.2 Discrete Least Squares Horn–Schunck Method 509
6.4.3 Horn–Schunck Algorithm 511
6.4.4 Lucas–Kanade Method 513
6.4.5 BBPW Method 517
6.4.6 Optical Flow Estimation for Affine Motion 521
6.4.7 Estimation of the Optical Flow for Large Displacements 523
6.4.8 Motion Estimation by Alignment 526
6.4.9 Motion Estimation with Techniques Based on Interest Points 535
6.4.10 Tracking Based on the Object Dynamics—Kalman Filter 542
6.5 Motion in Complex Scenes 556
6.5.1 Simple Method of Background Subtraction 557
6.5.2 BS Method with Mean or Median 558
6.5.3 BS Method Based on the Moving Gaussian Average . . . 559
6.5.4 Selective Background Subtraction Method 560
6.5.5 BS Method Based on Gaussian Mixture Model (GMM) 561
6.5.6 Background Modeling Using Statistical Method Kernel Density Estimation 563
6.5.7 Eigen Background Method 564
6.5.8 Additional Background Models 565
6.6 Analytical Structure of the Optical Flow of a Rigid Body 566
6.6.1 Motion Analysis from the Optical Flow Field 570
6.6.2 Calculation of Collision Time and Depth 573
6.6.3 FOE Calculation 577
6.6.4 Estimation of Motion Parameters for a Rigid Body 580
6.7 Structure from Motion 585
6.7.1 Image Projection Matrix 585
6.7.2 Methods of Structure from Motion 590
References 596

7 Camera Calibration and 3D Reconstruction 599
7.1 Introduction 599
7.2 Influence of the Optical System 600

7.3 Geometric Transformations Involved in Image Formation 602
7.4 Camera Calibration Methods 603
7.4.1 Tsai Method 605
7.4.2 Estimation of the Perspective Projection Matrix 610
7.4.3 Zhang Method 616
7.4.4 Stereo Camera Calibration 625
7.5 Stereo Vision and Epipolar Geometry 627
7.5.1 The Essential Matrix 629
7.5.2 The Fundamental Matrix 634
7.5.3 Estimation of the Essential and Fundamental Matrix . . . 638
7.5.4 Normalization of the 8-Point Algorithm 641
7.5.5 Decomposition of the Essential Matrix 642
7.5.6 Rectification of Stereo Images 645
7.5.7 3D Stereo Reconstruction by Triangulation 655
References 666

Index 669