
Texts in Computer Science

Series Editors

David Gries, Department of Computer Science, Cornell University, Ithaca, NY,
USA

Orit Hazzan, Faculty of Education in Technology and Science, Technion—Israel
Institute of Technology, Haifa, Israel

More information about this series at <http://www.springer.com/series/3191>

Dengsheng Zhang

Fundamentals of Image Data Mining

Analysis, Features, Classification
and Retrieval

Dengsheng Zhang
Federation University Australia
Churchill, VIC, Australia

ISSN 1868-0941 ISSN 1868-095X (electronic)
Texts in Computer Science
ISBN 978-3-030-17988-5 ISBN 978-3-030-17989-2 (eBook)
<https://doi.org/10.1007/978-3-030-17989-2>

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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 beloved wife Qin who makes this
possible*

Preface

Due to the rapid development of Internet and digital technology, a mammoth amount of digital data has been created in the world in just a few decades. The processing and management of big data including image data have become one of the great challenges facing humankind. Images are a dominant information source and communication method along with text. However, the processing and understanding of image data are far more difficult than dealing with textual data. Tremendous efforts have been made, and a large amount of research work has been carried out around the world in the past two decades to overcome the challenges of efficient management of image data. A significant progress has been achieved in the field of image data mining during this period of intensive research and experiments, highlighted by such breakthrough technologies as wavelets, MPEG, Google image search, convolutional neural network, machine learning, ImageNet, Matlab tool-boxes, etc.

Given the complexity of image data mining, there is a need for a deep analysis of and insight into the field, especially the latest development, to help researchers understand opportunities and challenges in the field. This book timely captures and presents cutting-edge techniques in the field of image data mining as well as foundational know-how for understanding them. This book provides a complete recipe for image data mining and is a treasure of techniques on image analysis/understanding, feature extraction, machine learning, and image retrieval. The book is built upon the author's career-long and high-impact research in the frontier of this exciting research field. Theories and concepts in the book are typically formulated with practical mathematical models which are realized using algorithms, real data from actual experiments, or working examples. Students and researchers in mathematics and the broader science disciplines will be able to use this book to understand the actual problems/applications in this field and gain firsthand experience of computing. Students and researchers in many areas of the computing discipline will be able to use this book to understand how fundamental and advanced maths are applied to solve the variety of computing problems.

Churchill, Australia

Dengsheng Zhang

About This Book

The book covers the complete know-how on image data mining including math tools, analysis, features, learning, and presentation. It has been organized into four parts: fundamentals, feature extraction, image classification, and image retrieval.

Part I of the book aims to equip readers with some essential tools for image mining. Specifically, Part I provides a brief and evolutionary journey from the classical Fourier transform (Chap. 1) to Gabor filters (Chap. 2) and to contemporary wavelet transform (Chap. 3). It prepares readers with fundamental math for some of the advanced mining techniques discussed in the book. Apart from the theories, this part also uses Fourier spectra, STFT spectrogram, Gabor filter spectra, and wavelet spectra to demonstrate how key information or features in an image can be captured by these fundamental transforms.

Parts II and III are the core of the book, which examine and analyze varieties of state-of-the-art models, tools, algorithms/procedures, and machines for image mining. In contrast to Part I which is mostly theoretical, these two parts focus on dealing with real image data and real image mining. Part II demonstrates how a variety of features can be mined or extracted from images for image representation; it covers three chapters which focus on color (Chap. 4), texture (Chap. 5) and shape (Chap. 6), respectively. Each chapter typically begins with simple methods or methods at the intro level and moves on to the more advanced methods in a natural flow. Most of the methods in Part II are demonstrated with intuitive illustrations.

If Part II is analog to raw mining, Part III is about refinery. Specifically, Part III presents readers with four powerful learning machines to classify image data, including Bayesian (Chap. 7), SVM (Chap. 8), ANN/CNN (Chap. 9) and DT (Chap. 10). Each chapter in this part begins with an icebreaking and introductory journey to give readers a big picture and an orientation to follow. It then navigates to the more advanced topics with illustrations to key concepts and components of the learning machine. The story of each machine learning method is typically told with concise maths, demonstrations, applications, and implementations.

After a breathtaking and arduous journey on image mining involving feature extraction and machine learning, readers are soothed with a recovering journey on image retrieval in Part IV. Part IV deals with putting images in order, inspecting the quality of a haul and organizing them for presentation or display. Indexing techniques suitable for image data are first described in detail in Chap. 11 followed by

the analysis of a number of image ranking techniques in Chap. 12. The part concludes in Chap. 13 with a number of interesting image presentation techniques and powerful image database visualization methods.

Key Features of the Book

A shortcut to AI. AI and machine learning are usually intimidating to many who don't have the sophisticated mathematics background. This book, however, offers readers a surprising shortcut to AI on machine learning by introducing four major machine learning tools with filtered and easy to understand mathematics using rich illustrations.

A natural marriage between maths and data. Maths and data can only be understood well when they are well matched. This book brings mathematics and computing into a single display and tells image stories with maths by a trained mathematician.

Visualization of image data mining. With more than 200 illustrations (multiple illustrations in some figures), it can be said that the book is a visualization of image data mining, making it very easy to read and understand for readers.

End of chapter summary. Every chapter of the book is equipped with an end of chapter summary to highlight the key points and connect the dots in the chapter.

Exercises. High-quality exercises with instructions or Matlab code have been created for most of the chapters in the book, giving readers the opportunities to test their skills learnt from the book.

Writing for scanning. The book makes extensive use of powerful techniques for scientific and academic writing including inverted pyramid writing, bullet lists, plain language, keyword headings, text chunking, analogy, scannable content, blurbs, etc. Due to writing for scanning, it makes reading the book very efficient and a good experience.

Contents

Part I Preliminaries

1	Fourier Transform	3
1.1	Introduction	3
1.2	Fourier Series	3
1.2.1	Sinusoids	3
1.2.2	Fourier Series	5
1.2.3	Complex Fourier Series	8
1.3	Fourier Transform	9
1.4	Discrete Fourier Transform	10
1.4.1	DFT	10
1.4.2	Uncertainty Principle	11
1.4.3	Nyquist Theorem	13
1.5	2D Fourier Transform	14
1.6	Properties of 2D Fourier Transform	14
1.6.1	Separability	14
1.6.2	Translation	15
1.6.3	Rotation	15
1.6.4	Scaling	18
1.6.5	Convolution Theorem	18
1.7	Techniques of Computing FT Spectrum	19
1.8	Summary	22
1.9	Exercises	22
	References	23
2	Windowed Fourier Transform	25
2.1	Introduction	25
2.2	Short-Time Fourier Transform	25
2.2.1	Spectrogram	27
2.3	Gabor Filters	28
2.3.1	Gabor Transform	28
2.3.2	Design of Gabor Filters	29
2.3.3	Spectra of Gabor Filters	31

2.4	Summary	32
2.5	Exercises	32
	References	34
3	Wavelet Transform	35
3.1	Discrete Wavelet Transform	35
3.2	Multiresolution Analysis	36
3.3	Fast Wavelet Transform	37
3.3.1	DTW Decomposition Tree	37
3.3.2	1D Haar Wavelet Transform	39
3.3.3	2D Haar Wavelet Transform	41
3.3.4	Application of DWT on Image	42
3.4	Summary	42
3.5	Exercises	44
 Part II Image Representation and Feature Extraction		
4	Color Feature Extraction	49
4.1	Introduction	49
4.2	Color Space	50
4.2.1	CIE XYZ, xyY Color Spaces	50
4.2.2	RGB Color Space	53
4.2.3	HSV, HSL and HSI Color Spaces	55
4.2.4	CIE LUV Color Space	60
4.2.5	$Y'CbCr$ Color Space	62
4.3	Image Clustering and Segmentation	63
4.3.1	K -Means Clustering	63
4.3.2	JSEG Segmentation	64
4.4	Color Feature Extraction	66
4.4.1	Color Histogram	66
4.4.1.1	Component Histogram	67
4.4.1.2	Indexed Color Histogram	68
4.4.1.3	Dominant Color Histogram	68
4.4.2	Color Structure Descriptor	69
4.4.3	Dominant Color Descriptor	71
4.4.4	Color Coherence Vector	72
4.4.5	Color Correlogram	74
4.4.6	Color Layout Descriptor	75
4.4.7	Scalable Color Descriptor	76
4.4.8	Color Moments	77
4.5	Summary	78
4.6	Exercises	79
	References	80

5	Texture Feature Extraction	81
5.1	Introduction	81
5.2	Spatial Texture Feature Extraction Methods	81
5.2.1	Tamura Textures	82
5.2.2	Gray Level Co-occurrence Matrices	84
5.2.3	Markov Random Field	86
5.2.4	Fractal Dimension	87
5.2.5	Discussions	88
5.3	Spectral Texture Feature Extraction Methods	89
5.3.1	DCT-Based Texture Feature	89
5.3.2	Texture Features Based on Gabor Filters	90
5.3.2.1	Gabor Filters	90
5.3.2.2	Gabor Spectrum	92
5.3.2.3	Texture Representation	93
5.3.2.4	Rotation-Invariant Gabor Features	94
5.3.3	Texture Features Based on Wavelet Transform	96
5.3.3.1	Selection and Application of Wavelets	96
5.3.3.2	Contrast of DWT and Other Spectral Transforms	100
5.3.3.3	Multiresolution Analysis	101
5.3.4	Texture Features Based on Curvelet Transform	102
5.3.4.1	Curvelet Transform	102
5.3.4.2	Discrete Curvelet Transform	104
5.3.4.3	Curvelet Spectra	107
5.3.4.4	Curvelet Features	107
5.3.5	Discussions	109
5.4	Summary	110
5.5	Exercises	110
	References	111
6	Shape Representation	113
6.1	Introduction	113
6.2	Perceptual Shape Descriptors	114
6.2.1	Circularity and Compactness	114
6.2.2	Eccentricity and Elongation	115
6.2.3	Convexity and Solidarity	116
6.2.4	Euler Number	117
6.2.5	Bending Energy	118
6.3	Contour-Based Shape Methods	118
6.3.1	Shape Signatures	119
6.3.1.1	Position Function	119
6.3.1.2	Centroid Distance	119
6.3.1.3	Angular Functions	121

	6.3.1.4	Curvature Signature	122
	6.3.1.5	Area Function	124
	6.3.1.6	Discussions	125
	6.3.2	Shape Context	125
	6.3.3	Boundary Moments	126
	6.3.4	Stochastic Method	127
	6.3.5	Scale Space Method	128
	6.3.5.1	Scale Space	128
	6.3.5.2	Curvature Scale Space	129
	6.3.6	Fourier Descriptor	130
	6.3.7	Discussions	132
	6.3.8	Syntactic Analysis	132
	6.3.9	Polygon Decomposition	133
	6.3.10	Chain Code Representation	135
	6.3.11	Smooth Curve Decomposition	135
	6.3.12	Discussions	136
6.4		Region-Based Shape Feature Extraction	136
	6.4.1	Geometric Moments	136
	6.4.2	Complex Moments	138
	6.4.3	Generic Fourier Descriptor	141
	6.4.4	Shape Matrix	144
	6.4.5	Shape Profiles	145
	6.4.5.1	Shape Projections	145
	6.4.5.2	Radon Transform	146
	6.4.6	Discussions	149
	6.4.7	Convex Hull	150
	6.4.8	Medial Axis	151
6.5		Summary	152
6.6		Exercises	153
		References	154

Part III Image Classification and Annotation

7	Bayesian Classification	161
7.1	Introduction	161
7.2	Naïve Bayesian Image Classification	164
	7.2.1 NB Formulation	164
	7.2.2 NB with Independent Features	166
	7.2.3 NB with Bag of Features	166
7.3	Image Annotation with Word Co-occurrence	166
7.4	Image Annotation with Joint Probability	169
7.5	Cross-Media Relevance Model	171
7.6	Image Annotation with Parametric Model	172

7.7	Image Classification with Gaussian Process	174
7.8	Summary	176
7.9	Exercises	177
	References	178
8	Support Vector Machine	179
8.1	Linear Classifier	179
8.1.1	A Theoretical Solution	180
8.1.2	An Optimal Solution	181
8.1.3	A Suboptimal Solution	182
8.2	K-Nearest Neighbors Classification	183
8.3	Support Vector Machine	184
8.3.1	The Perceptron	185
8.3.2	SVM—The Primal Form	186
8.3.2.1	The Margin Between Two Classes	186
8.3.2.2	Margin Maximization	189
8.3.2.3	The Primal Form of SVM	189
8.3.3	The Dual Form of SVM	190
8.3.3.1	The Dual Form Perceptron	191
8.3.4	Kernel-Based SVM	192
8.3.4.1	The Dual Form SVM Versus NN Classifier	192
8.3.4.2	Kernel Definition	193
8.3.4.3	Building New Kernels	196
8.3.4.4	The Kernel Trick	196
8.3.5	The Pyramid Match Kernel	198
8.3.6	Discussions	201
8.4	Fusion of SVMs	202
8.4.1	Fusion of Binary SVMs	202
8.4.2	Multilevel Fusion of SVMs	203
8.4.3	Fusion of SVMs with Different Features	203
8.5	Summary	204
	References	205
9	Artificial Neural Network	207
9.1	Introduction	207
9.2	Artificial Neurons	208
9.2.1	An AND Neuron	209
9.2.2	An OR Neuron	210
9.3	Perceptron	211
9.4	Nonlinear Neural Network	212
9.5	Activation and Inhibition	215
9.5.1	Sigmoid Activation	215
9.5.2	Shunting Inhibition	216

9.6	The Backpropagation Neural Network	217
9.6.1	The BP Network and Error Function	217
9.6.2	Layer K Weight Estimation and Updating	219
9.6.3	Layer $K - 1$ Weight Estimation and Updating	220
9.6.4	The BP Algorithm	222
9.7	Convolutional Neural Network	223
9.7.1	CNN Architecture	223
9.7.2	Input Layer	224
9.7.3	Convolution Layer 1 (C1)	224
9.7.3.1	2D Convolution	226
9.7.3.2	Stride and Padding	226
9.7.3.3	Bias	226
9.7.3.4	Volume Convolution in Layer C1	227
9.7.3.5	Depth of the Feature Map Volume	227
9.7.3.6	ReLU Activation	228
9.7.3.7	Batch Normalization	229
9.7.4	Pooling or Subsampling Layer 1 (S1)	230
9.7.5	Convolution Layer 2 (C2)	230
9.7.6	Hyperparameters	231
9.8	Implementation of CNN	232
9.8.1	CNN Architecture	232
9.8.2	Filters of the Convolution Layers	234
9.8.3	Filters of the Fully Connected Layers	234
9.8.4	Feature Maps of Convolution Layers	237
9.8.5	Matlab Implementation	239
9.9	Summary	241
	References	241
10	Image Annotation with Decision Tree	243
10.1	Introduction	243
10.2	ID3	245
10.2.1	ID3 Splitting Criterion	245
10.3	C4.5	247
10.3.1	C4.5 Splitting Criterion	247
10.4	CART	248
10.4.1	Classification Tree Splitting Criterion	248
10.4.2	Regression Tree Splitting Criterion	249
10.4.3	Application of Regression Tree	250
10.5	DT for Image Classification	252
10.5.1	Feature Discretization	252
10.5.2	Building the DT	254
10.5.3	Image Classification and Annotation with DT	255

10.6 Summary	258
References	259

Part IV Image Retrieval and Presentation

11 Image Indexing	263
11.1 Numerical Indexing	263
11.1.1 List Indexing	263
11.1.2 Tree Indexing	263
11.2 Inverted File Indexing	264
11.2.1 Inverted File for Textual Documents Indexing	265
11.2.2 Inverted File for Image Indexing	266
11.2.2.1 Determine the Area Weight aw	267
11.2.2.2 Determine the Position Weight pw	267
11.2.2.3 Determine the Relationship Weight rw	268
11.2.2.4 Inverted File for Image Indexing	269
11.3 Summary	269
References	270
12 Image Ranking	271
12.1 Introduction	271
12.2 Similarity Measures	271
12.2.1 Distance Metric	271
12.2.2 Minkowski-Form Distance	272
12.2.3 Mass-Based Distance	273
12.2.4 Cosine Distance	277
12.2.5 χ^2 Statistics	278
12.2.6 Histogram Intersection	278
12.2.7 Quadratic Distance	279
12.2.8 Mahalanobis Distance	280
12.3 Performance Measures	281
12.3.1 Recall and Precision Pair (RPP)	281
12.3.2 F -Measure	284
12.3.3 Percentage of Weighted Hits (PWH)	285
12.3.4 Percentage of Similarity Ranking (PSR)	285
12.3.5 Bullseye Accuracy	286
12.4 Summary	287
References	287
13 Image Presentation	289
13.1 Introduction	289
13.2 Caption Browsing	289
13.3 Category Browsing	290

13.3.1	Category Browsing on the Web	291
13.3.2	Hierarchical Category Browsing	291
13.4	Content Browsing	294
13.4.1	Content Browsing in 3D Space	295
13.4.2	Content Browsing with Focus	295
13.4.3	Force-Directed Content Browsing	295
13.5	Query by Example	297
13.6	Query by Keywords	301
13.7	Summary	304
	References	304
Appendix: Deriving the Conditional Probability of a Gaussian Process		305
Index		311

About the Author

Dr. Dengsheng Zhang received his Ph.D. in computing from Monash University, Australia, in 2002. He has also earned a Bachelor of Science in Mathematics in 1985 and a Bachelor of Arts in English in 1987. He is currently a Senior Lecturer at School of Science, Engineering and Information Technology in Federation University Australia. Dr. Zhang is also a guest professor of Xi'an University of Posts & Telecommunications.

Dr. Zhang has over 25 years of research experience in the field of artificial intelligence and big data analysis. His main research interests include pattern recognition, machine learning, multimedia data classification, and retrieval. He has published more than 100 refereed papers on international journals and conference proceedings in his career. His publications have so far attracted nearly 10,000 citations according to Google Scholar. One of his papers published in Pattern Recognition in 2007 has been awarded the Best Journal Article by the International Association of Pattern Recognition (IAPR) in CVPR2010.

Dr. Zhang has won two Australian Research Council (ARC) Discovery Grants to support his research on image and music data classification. He is also an experienced Ph.D. supervisor.

Dr. Zhang has over 30 years of international teaching experience and has lectured extensively on both undergraduate and postgraduate courses including mathematical analysis, business statistics, business programming, web design and development, multimedia development, computer systems, computer networks, e-commerce, computer interface design and development, mobile agent, business systems, general operations, professional development, computer models for business decisions and project management, etc.

List of Figures

Fig. 1.1	Fourier spectra of different images. a A scenic image at the left and its Fourier spectrum at the right; b a tree image in the left and its Fourier spectra in the right. The brighter the pixel, the higher magnitude of the spectrum.	4
Fig. 1.2	Three sine waves $\sin(nx)$ with different periods and frequencies.	4
Fig. 1.3	Three sine waves $\sin(2\pi nx)$ with different periods and frequencies.	6
Fig. 1.4	Illustration of $\int \sin x \sin(2x) dx = 0$	7
Fig. 1.5	Illustration of $\int \sin x \cos x dx = 0$	7
Fig. 1.6	Inverse relationship between spatial and frequency resolution. Left: the first 100 samples of the two sine waves; right: the last 100 samples of the two sine waves	12
Fig. 1.7	Illustrations of different sampling rates for three signals of the same time length	13
Fig. 1.8	FT spectra with different methods. a FT spectrum without scaling; b FT spectrum with thresholding value 10; c FT spectrum with linear scaling; d FT spectrum with log transform; e FT spectrum with both log and linear transform; and f FT spectrum with enhanced contrast from (e)	20
Fig. 1.9	FT spectra of different types of patterns. a Regular patterns and their FT spectra on the right; b random patterns and their FT spectra on the right; and c directional patterns and their FT spectra on the right	21
Fig. 2.1	Two images and their corresponding Fourier spectra on the right.	26
Fig. 2.2	Time–frequency illustration for FT and STFT	26
Fig. 2.3	The spectrogram of a sound wave	27
Fig. 2.4	The full width at half maximum of a Gaussian function.	29
Fig. 2.5	The half-amplitude of Gabor filters in the frequency domain using four scales and six orientations.	30
Fig. 2.6	Bandwidth tiling in frequency plane using Gaussian windows	30
Fig. 2.7	Spectra of Gabor filters at different scales and orientations.	31

Fig. 3.1	Different frequency tiling of spectral plane. Left: the wavelet spectrum; Right: the STFT spectrum.	37
Fig. 3.2	The 2D DWT decomposition tree for a lady image	38
Fig. 3.3	Illustration of 2D DWT process on a lady image. a Horizontal transform; b vertical transform; c spectrum of level 1 2D DWT decomposition.	38
Fig. 3.4	Two levels of 2D wavelet decomposition. a The spectrum plane of two levels of 2D DWT; b the spectrum of two levels of 2D DWT on the lady image; c the complete decomposition of the lady image	38
Fig. 3.5	Harr scaling function and wavelet function.	40
Fig. 3.6	Illustration of the computing process of Haar wavelet transform.	43
Fig. II.1	The 3D map for a 2D image.	46
Fig. 4.1	Visible light spectrum and the tristimulus.	50
Fig. 4.2	CIEXYZ color matching functions of human vision.	51
Fig. 4.3	Color matching function of spectral colors. Left: <i>rgb</i> curve of spectral colors; Right: projection of the <i>rgb</i> curve onto the 2D <i>rg</i> plane (cyan)	52
Fig. 4.4	CIE <i>xy</i> color gamut.	53
Fig. 4.5	The sRGB triangle gamut shown inside the CIE XYZ gamut	54
Fig. 4.6	RGB color space in 3D.	55
Fig. 4.7	a Artists' way of making reddish colors. b Components of HSV	56
Fig. 4.8	Hue and saturation. Left: pure colors on a ring; Right: hue-saturation wheel.	57
Fig. 4.9	HSV and HSL color spaces. a HSV cylinder; b HSL cylinder; c HSV cone; d HSL double cone	57
Fig. 4.10	A color image on the left and its H, S, V channels on the right columns	60
Fig. 4.11	MacAdam ellipses (magnified 10 times) on the CIE <i>xy</i> gamut	60
Fig. 4.12	CIE Luv and Lu'v'. Left: MacAdam ellipses on CIE <i>uv</i> gamut; Right: MacAdam ellipses on CIE <i>u'v'</i> gamut	61
Fig. 4.13	The flower image on the leftmost and its Y', Cb, and Cr channels on the right.	63
Fig. 4.14	An image segmentation using JSEG. a An original color image; b result of color quantization with 13 colors; c <i>J</i> -image at scale 3; d <i>J</i> -image at scale 2; e segmentation result at scale 3; f segmentation result at scale 2; g final result of segmentation after merging.	65

Fig. 4.15	RGB histograms of the Lena color image. Non-RGB colors are the areas of overlap between the R, G, B channels. Each channel is quantized into 256 colors or bins, which are on the horizontal axis, vertical axis shows the number of pixels in each bin or color.	66
Fig. 4.16	RGB channels of a color image. Top row: R, G, B channels of the flower image; Bottom row: corresponding gray level images of the R, G, B channels at the top row.	67
Fig. 4.17	Concatenation of histograms of individual R, G, B channels into a single histogram (216 bins, 72 bins for each channel)	68
Fig. 4.18	Indexed color histograms of a color image. a The flower image quantized with a global palette of 216 indexed colors (notice the visible distortion on the sky and grass); b indexed color histogram of (a) (216 bins); c the flower image quantized with an adaptive palette of 216 indexed colors; d dominant color histogram of (c)	69
Fig. 4.19	Computation of color structure descriptor. Left: a color image with a 3×3 structure moving through the image. The structure captures the red color 56 times; Right: a color image with a 3×3 structure moving through the image. The structure captures the red color 218 times.	70
Fig. 4.20	Accumulation of color structure descriptor. a A 5-color image and a 4×4 structuring element. b The accumulation of color structure histogram at a particular position of the structuring element in the image.	71
Fig. 4.21	Statistics of DCD numbers from 36,692 regions.	72
Fig. 4.22	Segmented regions and their dominant colors underneath. The dominant colors are shown according to their percentages in the region	73
Fig. 4.23	Removal of noisy colors from segmented regions. a Three sample regions; b dominant colors of corresponding regions above; c DCDs after discarding insignificant colors from (b)	73
Fig. 4.24	Computation of color correlogram. a A 4-color image; b the color correlogram of a for horizontal distance $k = 1$	75
Fig. 5.1	An example of computing directionality.	84
Fig. 5.2	Computation of a GLCM. An 8 gray levels image $I(x, y)$ on the left and its GLCM on the right	85
Fig. 5.3	FWHM sampling of spectral responses of Gabor filters in frequency plane	91
Fig. 5.4	Gabor filtered subbands for the lady image.	92

Fig. 5.5	Computation of Gabor texture descriptor. A straw image on the left and its energy map on the right. The higher the energy the brighter the block	93
Fig. 5.6	Computation of rotation-invariant Gabor texture descriptor. a A straw image; b energy map of (a) ; c a rotated image of (a) ; d energy map of (c) . The higher the energy, the brighter the block	95
Fig. 5.7	Frequency tiling of half frequency plan by Gabor filters, the ovals are the covered spectrum while the black holes are the lost spectrum.	96
Fig. 5.8	Mexican hat, Morlet, Daubechies, Meyer, Symlet 4, and Coiflet wavelets	97
Fig. 5.9	Wavelets in 3D space. Top left: Daubechies 2; Top right: Symlet 4; Bottom left: Coiflet 1; Bottom right: Mexican hat.	98
Fig. 5.10	Wavelet spectra for the Lena image	99
Fig. 5.11	Comparison of curvelet, wavelet and Gabor filter. a A curvelet; b a Daubechies wavelet and c a Gabor filter	103
Fig. 5.12	A curvelet and curvelet tiling in spatial domain. Left: a single curvelet with width 2^{-j} and length $2^{-j/2}$; Right: curvelets tuned to 2 scales at different orientations	103
Fig. 5.13	Edge representation using wavelets and curvelets	104
Fig. 5.14	Curvelet tiling of frequency plane with 5 level curvelets	104
Fig. 5.15	Curvelets at different scales are shown in the spatial domain (left) and in the frequency domain (right) respectively	105
Fig. 5.16	Curvelet subbands at different scales for a flower image (512×512). Each subband captures curvelet coefficients of the input image from one orientation	108
Fig. 5.17	Spectra of wavelets and Gabor filters for the flower image in Fig. 5.16a. a Wavelet spectra; b Gabor filters spectra.	108
Fig. 6.1	Examples of shape images	114
Fig. 6.2	Two different shapes with same circularity	115
Fig. 6.3	Computation of Eccentricity. a Eccentricity with principle axes; b eccentricity with minimum bounding box.	115
Fig. 6.4	A curled eel with 0 elongation	116
Fig. 6.5	A hand shape and its convex hull.	116
Fig. 6.6	Shapes with different Euler numbers.	117
Fig. 6.7	Examples of centroid distance signatures. Top row: a tree shape on the left and its centroid distance signature on the right; middle row: a ray fish on the left and its signature on the right; bottom row: an apple shape on the left and its signature on the right.	120

Fig. 6.8	Computation of chord length signature. a A sea snake shape on the left and its $r(t)$ function on the right; b another sea snake shape on the left and its $r(t)$ function on the right; c illustration of computing $r^*(t)$ at point P of a hammer shape.	121
Fig. 6.9	Computation of angular signatures. a A heart shape; b $\theta(t)$ of (a); c $ \theta(t) $ of (a); d $\varphi(t)$ of (a); e $\psi(t)$ of (a); f illustration of the computation of $\varphi(t)$	123
Fig. 6.10	Curvature signatures. Top: curvature signature of a tree shape from Fig. 6.7 without smoothing; Bottom: curvature signature of the same tree shape with a Gaussian smoothing.	124
Fig. 6.11	Computation of area signature. a Triangulation of a heart shape; b illustration on the calculation of the area of the shaded triangle	124
Fig. 6.12	Computation of shape context. a A point P on a shape boundary and all the vectors started from P ; b the log-polar histogram H of the vectors from P , the histogram H is the context of point P ; c the shape context map of shape of (a), each row of the context map is the flattened histogram of a point context, the number of rows is the number of sampled points	126
Fig. 6.13	Shape signature in scale space. a An original signature function $f(x)$ at the bottom and its successively smoothed versions on the top of it (up to scale 512), where σ is the scale of the smoothed function; b the interval tree derived from the zero-crossings of the second derivatives of the smoothed functions at the left, each (x, σ) in the interval tree corresponds to a zero-crossing point at position x and scale σ of the function at left-hand side.	128
Fig. 6.14	Computation of curvature scale space. A fish shape (left), its CSS contour map (center) and CSS peaks (right).	129
Fig. 6.15	The evolution of shape boundary as scale σ increases	130
Fig. 6.16	Reconstructed shapes of an apple shape using Fourier coefficients from a $r(t)$; b $z(t)$. In both (a) and (b), from left to right, the shapes are reconstructed using 5, 10, 20, 30, and all the Fourier coefficients, respectively	131
Fig. 6.17	Syntactic analysis of a hammer shape.	132
Fig. 6.18	Polygon approximation by merging	133
Fig. 6.19	Polygon approximation by splitting	134
Fig. 6.20	Computation of chain code. a Chain code in eight-connectivity; b chain code in four-connectivity	135
Fig. 6.21	Smooth curve decomposition. Left: a horse shape and its boundary segments; right: the calculation of the angle of a segment	136

Fig. 6.22	The first ten real Zernike polynomials	139
Fig. 6.23	The first 36 Zernike moments from order 1 to 10.	140
Fig. 6.24	Real parts of the ART basis functions	141
Fig. 6.25	Polar raster transform. a An original shape image in polar space; b polar raster sampled image of (a) plotted into Cartesian space	142
Fig. 6.26	Rotation invariant GFD. Top row: two shape images with different orientations; middle row: the polar raster sampled images of the two corresponding shapes at the top row; bottom row: the Fourier spectra images of the two corresponding images at the middle row.	143
Fig. 6.27	Computation of a shape matrix. A shape on the left and its shape matrix on the right.	144
Fig. 6.28	Computation of a polar raster shape matrix. Left: polar raster sampling of a shape; Right: its polar shape matrix	145
Fig. 6.29	Computation of shape profiles. A binary shape image with its vertical profile $P_v(x)$ (top) and horizontal profile $P_h(y)$ (right-hand side)	146
Fig. 6.30	A shape profile from Radon transform	147
Fig. 6.31	Shape profiles of a hammer image. Top row: A hammer shape and its Radon transforms at 90° and 8° ; Bottom row: the 90° and 8° profiles of the hammer shape after zooming in and realignment	148
Fig. 6.32	Radon transform of hammer shape. A hammer shape (left) and its Radon transform spectrum (right)	149
Fig. 6.33	A dog shape and its Radon transform spectrum	149
Fig. 6.34	Convex hull and concavity tree of an apple shape. a The convex hull of an apple shape and its concavities; b Concavity tree representation of the convex hull.	151
Fig. 6.35	Computation of media axis. Left: construction of medial axis of a rectangle shape using locus of circles; Right: the medial axis or skeleton of the hand shape	151
Fig. 6.36	Computation of a shape skeleton. a A horse shape image; b distance map of (a) ; c skeleton of the shape	152
Fig. III.1	An image to be classified into one of the classes	156
Fig. III.2	An image on the left and its color histogram on the right	156
Fig. III.3	Matching between an unknown image with a labeled image	157
Fig. III.4	Use a labeled image to identify all the mountain images in a database.	157
Fig. III.5	Sample apples (right) and the learnt apple model (left).	158
Fig. III.6	Computation of a generative model. Sample images are represented as features (right); a mixture distribution model is learnt from those sample features (left)	158

Fig. III.7	A machine is fitted between two classes of data	159
Fig. 7.1	Image classification with Naïve Bayesian method.	165
Fig. 7.2	Image annotation with co-occurrence of words.	168
Fig. 7.3	Association of semantic words with block features.	170
Fig. 7.4	Association of semantic words with region features	171
Fig. 7.5	Image annotation with parametric model	173
Fig. 7.6	Feature vectors shown as functions. Three histograms shown as vertical color bars and their respective functions shown as colored curves on the top	175
Fig. 7.7	A cluster of multidimensional data (green) and the approximation function of the data shown in pink	175
Fig. 8.1	Comparison between NN and K-NN. a The data to be classified; b classification result from a 1-NN classifier; c classification result from a 5-NN classifier.	184
Fig. 8.2	Hyperplanes between two classes of data	186
Fig. 8.3	The optimal hyperplane H between two classes of data	187
Fig. 8.4	Two classes of data and the hyperplanes separating them.	188
Fig. 8.5	Mapping of nonlinear data to linear data in higher dimensional space. a An original nonlinear data in 2D space; b transformed data in 3D space using a quadratic mapping function ϕ	197
Fig. 8.6	Computation of pyramid match kernel. An image is divided into three levels of grids. At each level of the grid, a histogram is computed for each block of the grid. Histograms at each level are given a weight and the weighted histograms are then concatenated as a feature vector	200
Fig. 8.7	Histogram intersection of two normal distributions.	201
Fig. 8.8	Approximation of class boundary using piecewise hyperplanes.	202
Fig. 8.9	A fusion of binary SVM classifiers.	202
Fig. 8.10	A 3 levels fusion of SVMs	203
Fig. 8.11	A 3 levels class-by-class fusion of SVMs with both global and local features.	204
Fig. 9.1	A neuron of human brain	208
Fig. 9.2	The alignment of an artificial neuron with a biologic neuron.	208
Fig. 9.3	The modeling of an artificial neuron.	209
Fig. 9.4	The data of AND can be separated by a single line $-1.5 + x_1 + x_2 = 0$	210
Fig. 9.5	The neuron which implements the AND function.	210
Fig. 9.6	The data of OR can be separated by a single line $-0.5 + x_1 + x_2 = 0$	211
Fig. 9.7	The neuron which implements the OR function	211

Fig. 9.8	Outputs of an XOR cannot be separated by a single line in 2D space.	213
Fig. 9.9	The convex data region in the center cannot be separated by a single line in 2D space	213
Fig. 9.10	A three-layer neural network to implement XOR. The two linear classifiers at the hidden layer are ANDed at the third layer	214
Fig. 9.11	A three-layer neural network with a hidden layer which can classify generic convex nonlinear data	214
Fig. 9.12	A non-convex data region (black dots) is approximated by three convex regions	214
Fig. 9.13	A four-layer neural network which can classify any nonlinear data	215
Fig. 9.14	Sigmoid function	215
Fig. 9.15	A shunting inhibitory neuron	216
Fig. 9.16	A K -layer backpropagation neural network.	218
Fig. 9.17	Illustration of a connection between node i in layer $K - 1$ and node j in layer K in a K -layer BP neural network	219
Fig. 9.18	Illustration of all connections relevant to w_{mi}^{K-1} in a K -layer BP neural network. Relevant connections are shown in red lines	221
Fig. 9.19	Architecture of a CNN	224
Fig. 9.20	Volume convolution. a Demonstration of volume filter and high-dimensional convolution; b demonstration of high-dimensional convolution $\mathbf{x} * w^0$ which can be done using a series of 2D convolutions. Each of the filter channels $a, b, c,$ and d is convoluted with each of the corresponding image channels $A, B, C,$ and D	225
Fig. 9.21	2D convolution. An image \mathbf{x} is convoluted with a filter window w and the result of the convolution is given by $\mathbf{x} * w$ at the right-hand side.	226
Fig. 9.22	An input image is convoluted with two volume filters w^0 and w^1 . The two result feature maps are at the rightmost hand side (light red). The <i>stride</i> value is 2 and there is no padding	228
Fig. 9.23	Illustration of a max-pooling. The maximum value of each quarter block of the left image is computed as the output value of the max-pooling image at the right-hand side.	230
Fig. 9.24	Volume convolution in layer C2. Volume filter w^0 is convoluted with feature maps A–D while volume filter w^1 is convoluted with feature maps C–F	231
Fig. 9.25	Pretrained filters of 6 of the 13 VGG16 convolution layers	235
Fig. 9.26	Filters of the fully connected layers	236

Fig. 9.27	Feature maps from different convolution layers of VGG16.	238
Fig. 9.28	Channels from each of the five blocks of VGG16 net	239
Fig. 10.1	A DT for image classification.	244
Fig. 10.2	Contrast between linear regression and regression tree	251
Fig. 10.3	A regression tree for predicting median house price ('000) in California from the geographic coordinates of Fig. 10.4. Legend: La = latitude, Lo = longitude, y = yes, n = no	251
Fig. 10.4	Data map of actual median house prices in California and the tree partition of the data map, the darker the color, the higher the house value	252
Fig. 10.5	A DT for image classification without post-pruning, CT = color and texture	256
Fig. 10.6	The DT from Fig. 10.5 after pruning	258
Fig. 11.1	A k - d tree for a 2D dataset of 10 data	264
Fig. 11.2	An image with labeled regions	265
Fig. 11.3	A flower in the center with 4 segmented regions	267
Fig. 11.4	Calculation of d and d_{max} for a animal, b sky and c grass regions	268
Fig. 12.1	Unit circles of Minkowski distance with different p values. a $p = \frac{1}{2}$; b $p = 1$; c $p = 2$	273
Fig. 12.2	Top: a histogram feature vector; Bottom: the log-transformed histogram feature vector from the top histogram.	274
Fig. 12.3	The three images have the same histogram.	274
Fig. 12.4	The two images with different brightness have almost completely different histograms.	275
Fig. 12.5	Illustration of m_p dimension calculation between two data points x and y	276
Fig. 12.6	Comparison between the cosine distance and L_p distance. a $L_2(\mathbf{x}, \mathbf{y}) = L_2$; b $\cos(\mathbf{x}, \mathbf{y}) = \cos\theta$; c $L_1(\mathbf{x}, \mathbf{y}) = d_{11} + d_{12}$	277
Fig. 12.7	Histogram intersection of two histograms shown as gray area.	278
Fig. 12.8	An RPP curve from an actual image retrieval.	283
Fig. 12.9	The RPP curve and F curve from an actual image retrieval.	284
Fig. 12.10	The F curves for two different RPP curves with the same AUC.	285
Fig. 13.1	Browsing images using image captions by Windows File Explorer	290
Fig. 13.2	Category browsing using MS Windows File Explorer	291
Fig. 13.3	A website of category browsing	292
Fig. 13.4	Hierarchical organization of image categories on a website	293
Fig. 13.5	Image presentation based on content browsing	294
Fig. 13.6	Content browsing using a cylindrical image map	296
Fig. 13.7	Content browsing using a spherical image map	296

Fig. 13.8	Content browsing using a curved wall image map	297
Fig. 13.9	Content browsing with focus and spiral structure	298
Fig. 13.10	Force-directed content browsing	299
Fig. 13.11	A QBE example from Google	301
Fig. 13.12	Query images for Google’s “Search by image”.	301
Fig. 13.13	A content-based QBK system.	302
Fig. 13.14	A QBK example from Google. a An example from Google’s Image search; b category browsing from Google’s Image search	303
Fig. 13.15	A hierarchical QBK in Google. An enlarged image at the black strip and its relevant category at the right-hand side of the enlarged image	303

List of Tables

Table 6.1	List of Zernike moments up to order 10.	140
Table 9.1	AND gate.	210
Table 11.1	A conceptual inverted file for textual document indexing	265
Table 11.2	An inverted file after sorting	266
Table 11.3	A conceptual inverted file for image document indexing	269
Table 11.4	The simplified inverted file from Table 11.3 after sorting	269