

Advances in Pattern Recognition

For further volumes:

<http://www.springer.com/series/4205>

Nanning Zheng · Jianru Xue

Statistical Learning and Pattern Analysis for Image and Video Processing

 Springer

Prof. Nanning Zheng
Xi'an Jiaotong University
Inst. Artificial Intelligence &
Robotics
28 Xianningxilu, Xi'an, 710049
China, People's Republic
nnzheng@mail.xjtu.edu.cn

Prof. Jianru Xue
Xi'an Jiaotong University
Inst. Artificial Intelligence &
Robotics
28 Xianningxilu, Xi'an, 710049
China, People's Republic
jrxue@mail.xjtu.edu.cn

Series editor

Professor Sameer Singh, PhD
Research School of Informatics
Loughborough University
Loughborough, UK

ISSN 1617-7916
ISBN 978-1-84882-311-2 e-ISBN 978-1-84882-312-9
DOI 10.1007/978-1-84882-312-9
Springer Dordrecht Heidelberg London New York

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2009932139

© Springer-Verlag London Limited 2009

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licenses issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

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

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Why are We Writing This Book?

Visual data (graphical, image, video, and visualized data) affect every aspect of modern society. The cheap collection, storage, and transmission of vast amounts of visual data have revolutionized the practice of science, technology, and business. Innovations from various disciplines have been developed and applied to the task of designing intelligent machines that can automatically detect and exploit useful regularities (patterns) in visual data. One such approach to machine intelligence is statistical learning and pattern analysis for visual data.

Over the past two decades, rapid advances have been made throughout the field of visual pattern analysis. Some fundamental problems, including perceptual grouping, image segmentation, stereo matching, object detection and recognition, and motion analysis and visual tracking, have become hot research topics and test beds in multiple areas of specialization, including mathematics, neuron-biometry, and cognition. A great diversity of models and algorithms stemming from these disciplines has been proposed. To address the issues of ill-posed problems and uncertainties in visual pattern modeling and computing, researchers have developed rich toolkits based on pattern analysis theory, harmonic analysis and partial differential equations, geometry and group theory, graph matching, and graph grammars.

Among these technologies involved in intelligent visual information processing, statistical learning and pattern analysis is undoubtedly the most popular and important approach, and it is also one of the most rapidly developing fields, with many achievements in recent years. Above all, it provides a unifying theoretical framework for intelligent visual information processing applications.

The main topics of this book are the modeling and computing of visual patterns in image sequences and the methods required to construct intelligent video analysis systems. For visual pattern modeling, we apply statistical learning and statistical pattern analysis to extract semantic visual objects. In our view, such models can be learned efficiently to emulate complex scenes in the image sequence. For video analysis system building, the methods presented here are based on techniques of statistical computing, such as motion analysis, inferring underlying states of objects

of interest, and reducing the dimensionality of video data while preserving useful information as much as possible.

How the Book is Organized

This book provides a comprehensive overview of theories, methodologies, and recent developments in the field of statistical learning and statistical analysis for visual pattern modeling and computing. We had three objectives in selecting topics to cover. We wish to 1) describe a solid theoretical foundation, 2) provide a comprehensive summary of the latest advances of recent years, and 3) present typical issues to be considered in making a real system for visual information processing. We have tried to achieve a balance between these three objectives. The rest of this book is organized as follows:

Chapter 1 is devoted to constructing the theoretic basis for pattern analysis and statistical learning. The fundamentals of statistical pattern recognition and statistical learning are presented via introducing the general framework of a statistical pattern recognition system. We also discuss pattern representation and classification, two important components of such a system, as well as concepts involved in three main approaches to statistical learning: supervised learning, semistatistical learning, and unsupervised learning. This introduction leads to the development of three parts of the whole book.

In the first part, we focus on the unsupervised learning of visual pattern representational models for objects in images, which covers through Chapters 2 to 5. Usually, what a vision algorithm can accomplish depends crucially on how much it knows about content of the visual scenes. This knowledge can be mathematically represented by simple but general models that can realistically characterize visual patterns in the ensemble of visual data. Representation and computation are thus two principal problems in visual computing. We provide a comprehensive survey of recent advances in statistical learning and pattern analysis with respect to these two problems. Chapter 2 discusses cluster analysis and perceptual grouping algorithms used in unsupervised visual pattern analysis. The systematic approaches for deriving these models are also illustrated step by step. Chapters 3 through 5 focus on representing and learning visual patterns in both spatial and temporal domains. Chapter 3 describes component analysis approaches, which are used to find hidden components via visual data analysis techniques. Chapter 4 discusses the manifold learning perspective on visual pattern representation, dimensionality reduction, and classification problems. Chapter 5 presents a review of recent advances in the adaptive wavelet transform for image and video coding.

In the second part, we introduce the supervised learning of visual patterns in images, which is covered in Chapter 6. We focus on supervised statistical pattern analysis and introduce concepts and major techniques in feature extraction and selection as well as classifier design. Especially, we introduce statistical machine learning techniques by examining the support vector machine and AdaBoost classifier.

In the third part, we focus on the visual pattern analysis in video, which covers through Chapters 7 to 11. In this part, we discuss methodologies for building intelligent video analysis systems such as object detection, tracking, and recognition in video. Chapter 7 focuses on the critical aspects of motion analysis, including statistical optical flow, model-based motion analysis, and joint motion estimation and segmentation. For the object-level motion analysis, we first introduce the sequential Bayesian estimation framework in Chapter 8, which acts as the theoretic basis for visual tracking, and then present approaches to constructing a representation model of specific objects. Then, in Chapter 9, we present a probabilistic fusion framework for robust tracking. Chapters 10 and 11 offer a multitarget tracking in video (MTTV) formulation that exploits a Markov network whose solution is arrived at using Monte Carlo-based belief propagation. Using this approach, problems including occlusion and various number of objects in MTTV are addressed.

Finally, in Chapter 12, we make an in-depth discussion of visual data processing in the cognitive process. A new scheme of association memory and new architecture of artificial intelligent system with attractors of chaos are also addressed. We argue that to make a breakthrough in current research on intelligent visual data processing, people should pay great attention to the mechanism of cognition and selective attention.

Acknowledgments

First, we thank the National Science Foundation of China (Grant Nos. 608750081 and 60635050) for their continuing support of our research in the image and video processing field. Our interest in statistical learning and pattern analysis approaches to image and video processing could not have developed to the point of writing this book without their funding.

We also thank Xi'an Jiaotong University, for their support of our efforts in writing this book, and Wayne Wheeler, Senior editor of Springer-Verlag, for his suggestion in writing this book. Another major source of stimulation for this effort comes from our former graduate students, who have provided a continuing stream of new ideas and insights. Those who have made contributions directly reflected in this book include Mr. Xiaoping Zhong, Dr. Weixiang Liu, and Dr. Zaide Liu.

A special thanks goes to Dr. Clayton McMillan, who read the first draft very carefully and made numerous corrections for improvements. We also thank Dr. Shaoyi Du and Dr. Gaofeng Meng for their dedication and patience in preparing drafts and writing Chapter 6. The second author would like to thank Prof. Songchun Zhu for the hosting of his visiting in the Center for Image and Vision Science in University of California, Los Angeles, when he was writing this book.

Last but not least, we are greatly indebted to our families: our parents and our wives and children for their unconditional support, encouragement, and continuing love throughout our research work.

Contents

- 1 Pattern Analysis and Statistical Learning 1**
 - 1.1 Introduction 1
 - 1.1.1 Statistical Pattern Recognition 2
 - 1.1.2 Pattern Theory 4
 - 1.2 Statistical Classification 6
 - 1.2.1 Feature Extraction and Selection 6
 - 1.2.2 Classifier 7
 - 1.3 Visual Pattern Representation 8
 - 1.3.1 The Curse of Dimensionality 9
 - 1.3.2 Dimensionality Reduction Techniques 9
 - 1.4 Statistical Learning 10
 - 1.4.1 Prediction Risk 11
 - 1.4.2 Supervised, Unsupervised, and Others 12
 - 1.5 Summary 14
 - References 14
- 2 Unsupervised Learning for Visual Pattern Analysis 15**
 - 2.1 Introduction 15
 - 2.1.1 Unsupervised Learning 15
 - 2.1.2 Visual Pattern Analysis 16
 - 2.1.3 Outline 17
 - 2.2 Cluster Analysis 17
 - 2.3 Clustering Algorithms 21
 - 2.3.1 Partitional Clustering 21
 - 2.3.2 Hierarchical Clustering 30
 - 2.4 Perceptual Grouping 33
 - 2.4.1 Hierarchical Perceptual Grouping 33
 - 2.4.2 Gestalt Grouping Principles 35
 - 2.4.3 Contour Grouping 39
 - 2.4.4 Region Grouping 45
 - 2.5 Learning Representational Models for Visual Patterns 47

2.6	Summary	48
	Appendix	48
	References	48
3	Component Analysis	51
3.1	Introduction	51
3.2	Overview of Component Analysis	54
3.3	Generative Models	55
3.3.1	Principal Component Analysis	55
3.3.2	Nonnegative Matrix Factorization	66
3.3.3	Independent Component Analysis	72
3.4	Discriminative Models	76
3.4.1	Linear Discriminative Analysis	76
3.4.2	Oriented Component Analysis	79
3.4.3	Canonical Correlation Analysis	79
3.4.4	Relevant Component Analysis	81
3.5	Standard Extensions of the Linear Model	83
3.5.1	Latent Variable Analysis	83
3.5.2	Kernel Method	83
3.6	Summary	83
	References	84
4	Manifold Learning	87
4.1	Introduction	87
4.2	Mathematical Preliminaries	91
4.2.1	Manifold Related Terminologies	91
4.2.2	Graph Related Terminologies	92
4.3	Global Methods	94
4.3.1	Multidimensional Scaling	94
4.3.2	Isometric Feature Mapping	95
4.3.3	Variants of the Isomap	96
4.4	Local Methods	100
4.4.1	Locally Linear Embedding	100
4.4.2	Laplacian Eigenmaps	103
4.4.3	Hessian Eigenmaps	107
4.4.4	Diffusion Maps	109
4.5	Hybrid Methods: Global Alignment of Local Models	113
4.5.1	Global Coordination of Local Linear Models	113
4.5.2	Charting a Manifold	115
4.5.3	Local Tangent Space Alignment	117
4.6	Summary	117
	Appendix	118
	References	118

5	Functional Approximation	121
5.1	Introduction	121
5.2	Modeling and Approximating the Visual Data	124
5.2.1	On Statistical Analysis	125
5.2.2	On Harmonic Analysis	126
5.2.3	Issues of Approximation and Compression	127
5.3	Wavelet Transform and Lifting Scheme	129
5.3.1	Wavelet Transform	129
5.3.2	Constructing a Wavelet Filter Bank	130
5.3.3	Lifting Scheme	132
5.3.4	Lifting-Based Integer Wavelet Transform	133
5.4	Optimal Integer Wavelet Transform	134
5.5	Introducing Adaptability into the Wavelet Transform	136
5.5.1	Curve Singularities in an Image	137
5.5.2	Anisotropic Basis	137
5.5.3	Adaptive Lifting-Based Wavelet	139
5.6	Adaptive Lifting Structure	140
5.6.1	Adaptive Prediction Filters	140
5.6.2	Adaptive Update Filters	142
5.7	Adaptive Directional Lifting Scheme	143
5.7.1	ADL Framework	144
5.7.2	Implementation of ADL	145
5.8	Motion Compensation Temporal Filtering in Video Coding	148
5.8.1	Overview of MCTF	148
5.8.2	MC in MCTF	151
5.8.3	Adaptive Lifting-Based Wavelets in MCTF	152
5.9	Summary and Discussions	153
	References	154
6	Supervised Learning for Visual Pattern Classification	159
6.1	Introduction	159
6.2	An Example of Supervised Learning	160
6.3	Support Vector Machine	163
6.3.1	Optimal Separating Hyper-plane	163
6.3.2	Realization of SVM	167
6.3.3	Kernel Function	169
6.4	Boosting Algorithm	171
6.4.1	AdaBoost Algorithm	172
6.4.2	Theoretical Analysis of AdaBoost	173
6.4.3	AdaBoost Algorithm as an Additive Model	176
6.5	Summary	178
	Appendix	178
	References	179

7	Statistical Motion Analysis	181
7.1	Introduction	181
7.1.1	Problem Formulation	181
7.1.2	Overview of Computing Techniques	183
7.2	Bayesian Estimation of Optical Flow	186
7.2.1	Problem Formulation	186
7.2.2	MAP Estimation	190
7.2.3	Occlusion	192
7.3	Model-Based Motion Analysis	193
7.3.1	Motion Models	194
7.3.2	Statistical Model Selection	195
7.3.3	Learning Parameterized Models	196
7.4	Motion Segmentation	201
7.4.1	Layered Model: Multiple Motion Models	202
7.4.2	Clustering Optical Flow Field into Layers	204
7.4.3	Mixture Estimation for Layer Extraction	205
7.5	Statistics of Optical Flow	208
7.5.1	Statistics of Optical Flow	208
7.5.2	Motion Prior Modeling	210
7.5.3	Contrastive Divergence Learning	211
7.6	Summary	212
	Appendix	212
	References	214
8	Bayesian Tracking of Visual Objects	217
8.1	Introduction	217
8.2	Sequential Bayesian Estimation	219
8.2.1	Problem Formulation of Bayesian Tracking	220
8.2.2	Kalman Filter	221
8.2.3	Grid-Based Methods	222
8.2.4	Sub-optimal Filter	222
8.3	Monte Carlo Filtering	224
8.3.1	Problem Formulation	224
8.3.2	Sequential Importance Sampling	226
8.3.3	Sequential Monte Carlo Filtering	231
8.3.4	Particle Filter	232
8.4	Object Representation Model	235
8.4.1	Visual Learning for Object Representation	236
8.4.2	Active Contour	237
8.4.3	Appearance Model	241
8.5	Summary	243
	References	244

9	Probabilistic Data Fusion for Robust Visual Tracking	245
9.1	Introduction	245
9.2	Earlier Work on Robust Visual Tracking	248
9.3	Data Fusion-Based Visual Tracker	251
9.3.1	Sequential Bayesian Estimator	251
9.3.2	The Four-Layer Data Fusion Visual Tracker	253
9.4	Layer 1: Visual Cue Fusion	255
9.4.1	Fusion Rules: Product Versus Weighted Sum	255
9.4.2	Adaptive Fusion Rule	257
9.4.3	Online Approach to Determining the Reliability of a Visual cue	258
9.5	Layer 2: Model Fusion	260
9.5.1	Pseudo-Measurement-Based Multiple Model Method	261
9.5.2	Likelihood Function	263
9.6	Layer 3: Tracker Fusion	264
9.6.1	Problem Formulation	265
9.6.2	Interactive Multiple Trackers	266
9.6.3	Practical Issues	267
9.7	Sensor Fusion	269
9.8	Implementation Issues and Empirical Results	271
9.8.1	Visual Cue Fusion Layer	271
9.8.2	Model Fusion Layer	274
9.8.3	Tracker Fusion Layer	276
9.8.4	Bottom-Up Fusion with a Three-Layer Structure	281
9.8.5	Multi-Sensor Fusion Tracking System Validation	281
9.9	Summary	283
	References	284
10	Multitarget Tracking in Video-Part I	287
10.1	Introduction	287
10.2	Overview of MTTV Methods	290
10.3	Static Model for Multitarget	292
10.3.1	Problem formulation	292
10.3.2	Observation Likelihood Function	294
10.3.3	Prior Model	295
10.4	Approximate Inference	296
10.4.1	Model Approximation	296
10.4.2	Algorithm Approximation	299
10.5	Fusing Information from Temporal and Bottom-Up Detectors	302
10.6	Experiments and Discussions	304
10.6.1	Proof-of-Concept	305
10.6.2	Comparison with Other Trackers	308
10.6.3	The Efficiency of the Gibbs Sampler	315
10.7	Summary	315
	References	315

11	Multi-Target Tracking in Video – Part II	319
11.1	Introduction	319
11.2	Overview of the MTTV Data Association Mechanism	322
11.2.1	Handling Data Association Explicitly	322
11.2.2	Handling Data Association Implicitly	324
11.2.3	Detection and Tracking	325
11.3	The Generative Model for MTT	326
11.3.1	Problem Formulation	326
11.3.2	The Generative Model	327
11.4	Approximating The Marginal Term	329
11.4.1	The State Prediction	330
11.4.2	Existence and Association Posterior	332
11.5	Approximating the Interactive Term	334
11.6	Hybrid Measurement Process	335
11.7	Experiments and Discussion	335
11.7.1	Tracking Soccer Players	336
11.7.2	Tracking Pedestrians in a Dynamic Scene	337
11.7.3	Discussion	337
11.8	Summary	340
	References	340
12	Information Processing in Cognition Process and New Artificial Intelligent Systems	343
12.1	Introduction	343
12.2	Cognitive Model: A Prototype of Intelligent System	345
12.3	Issues in Theories and Methodologies of Current Brain Research and Vision Science	347
12.4	Interactive Behaviors and Selective Attention in the Process of Visual Cognition	351
12.5	Intelligent Information Processing and Modeling Based on Cognitive Mechanisms	353
12.5.1	Cognitive Modeling and Behavioral Control in Complex Systems in an Information Environment	353
12.5.2	Distributed Cognition	356
12.5.3	Neurophysiological Mechanism of Learning and Memory and Information Processing Model	358
12.6	Cognitive Neurosciences and Computational Neuroscience	359
12.6.1	Consciousness and Intention Reading	360
12.6.2	The Core of Computational Neuroscience Is to Compute and Interpret the States of Nervous System	360
12.7	Soft Computing Method	360
12.8	Summary	361
	References	362
	Index	363