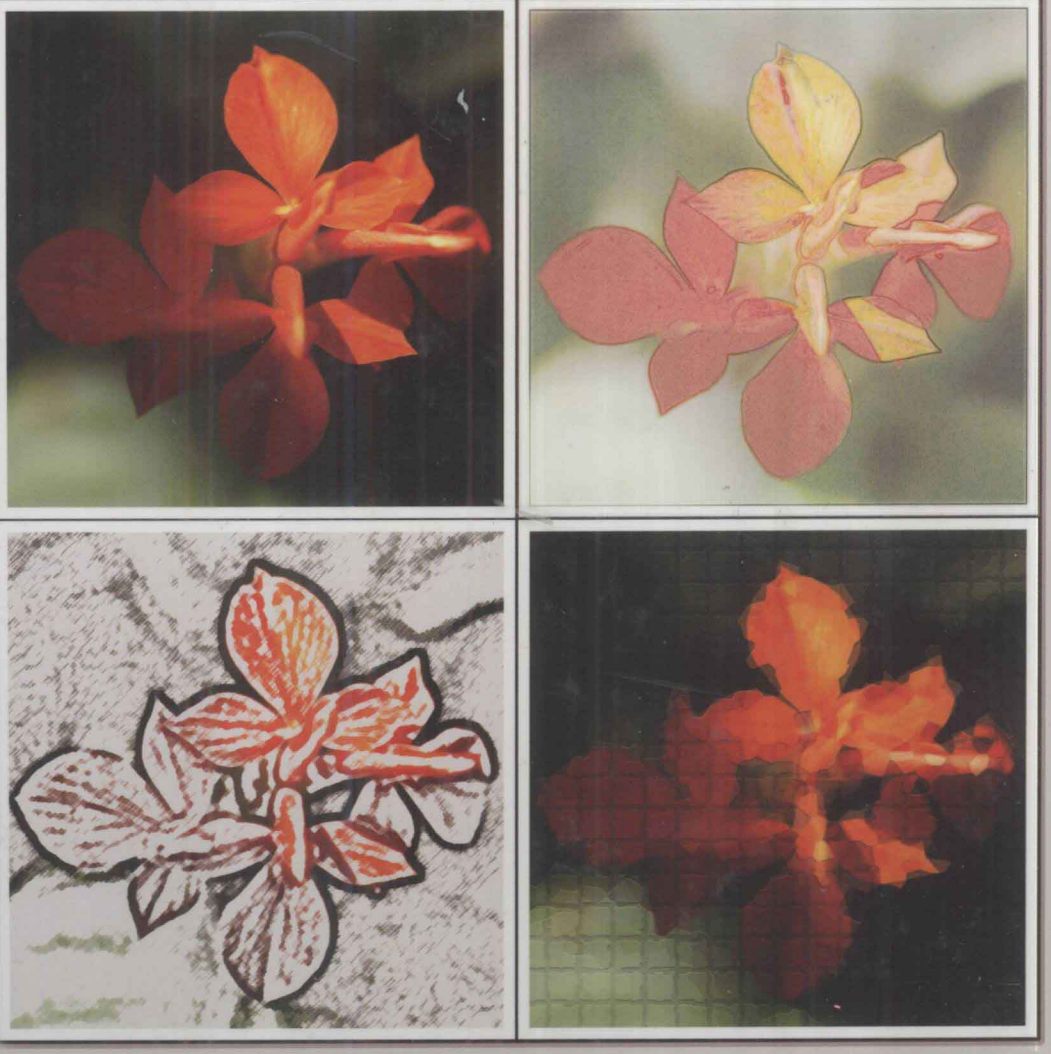
Chapman & Hall/CRC
Data Mining and Knowledge Discovery Series

Zhongfei Zhang Ruofei Zhang

Multimedia Data Mining

A Systematic Introduction to Concepts and Theory



CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Chapman & Hall/CRC
Data Mining and Knowledge Discovery Series

Multimedia Data Mining

A Systematic Introduction to Concepts and Theory

Zhongfei Zhang
Ruofei Zhang

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

The cover images were provided by Yu He, who also participated in the design of the cover page.

Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2009 by Taylor & Francis Group, LLC
Chapman & Hall/CRC is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-58488-966-3 (Hardcover)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Zhang, Zhongfei.
Multimedia data mining : a systematic introduction to concepts and theory / Zhongfei Zhang, Ruofei Zhang.
p. cm. -- (Chapman & Hall/CRC data mining and knowledge discovery series)
Includes bibliographical references and index.
ISBN 978-1-58488-966-3 (hardcover : alk. paper)
1. Multimedia systems. 2. Data mining. I. Zhang, Ruofei. II. Title. III. Series.

QA76.575.Z53 2008
006.7--dc22 2008039398

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Multimedia Data Mining

A Systematic Introduction
to Concepts and Theory

Chapman & Hall/CRC
Data Mining and Knowledge Discovery Series

SERIES EDITOR

Vipin Kumar
University of Minnesota
Department of Computer Science and Engineering
Minneapolis, Minnesota, U.S.A

AIMS AND SCOPE

This series aims to capture new developments and applications in data mining and knowledge discovery, while summarizing the computational tools and techniques useful in data analysis. This series encourages the integration of mathematical, statistical, and computational methods and techniques through the publication of a broad range of textbooks, reference works, and handbooks. The inclusion of concrete examples and applications is highly encouraged. The scope of the series includes, but is not limited to, titles in the areas of data mining and knowledge discovery methods and applications, modeling, algorithms, theory and foundations, data and knowledge visualization, data mining systems and tools, and privacy and security issues.

PUBLISHED TITLES

UNDERSTANDING COMPLEX DATASETS: Data Mining with Matrix Decompositions
David Skillicorn

COMPUTATIONAL METHODS OF FEATURE SELECTION
Huan Liu and Hiroshi Motoda

CONSTRAINED CLUSTERING: Advances in Algorithms, Theory, and Applications
Sugato Basu, Ian Davidson, and Kiri L. Wagstaff

KNOWLEDGE DISCOVERY FOR COUNTERTERRORISM AND LAW ENFORCEMENT
David Skillicorn

MULTIMEDIA DATA MINING: A Systematic Introduction to Concepts and Theory
Zhongfei Zhang and Ruofei Zhang

试读结束：需要全本请在线购买： www.ertongbook.com

To my parents, Yukun Zhang and Ming Song; my sister, Xuefei; and my sons, Henry and Andrew
Zhongfei (Mark) Zhang

To my parents, sister, and wife for their support and tolerance
Ruofei Zhang

Foreword

I am delighted to introduce the first book on multimedia data mining. When I came to know about this book project undertaken by two of the most active young researchers in the field, I was pleased that this book is coming in an early stage of a field that will need it more than most fields do. In most emerging research fields, a book can play a significant role in bringing some maturity to the field. Research fields advance through research papers. In research papers, however, only a limited perspective can be provided about the field, its application potential, and the techniques required and already developed in the field. A book gives such a chance. I liked the idea that there will be a book that will try to unify the field by bringing in disparate topics already available in several papers that are not easy to find and understand. I was supportive of this book project even before I had seen any material on it. The project was a brilliant and a bold idea by two active researchers. Now that I have it on my screen, it appears to be even a better idea.

Multimedia started gaining recognition in the 1990s as a field. Processing, storage, communication, and capture and display technologies had advanced enough that researchers and technologists started building approaches to combine information in multiple types of signals such as audio, images, video, and text. Multimedia computing and communication techniques recognize correlated information in multiple sources as well as insufficiency of information in any individual source. By properly selecting sources to provide complementary information, such systems aspire, much like the human perception system, to create a holistic picture of a situation using only partial information from separate sources.

Data mining is a direct outgrowth of progress in data storage and processing speeds. When it became possible to store large volumes of data and run different statistical computations to explore all possible and even unlikely correlations among data, the field of data mining was born. Data mining allowed people to hypothesize relationships among data entities and explore support for those. This field has been applied to applications in many diverse domains and keeps getting more applications. In fact, many new fields are a direct outgrowth of data mining, and it is likely to become a powerful computational tool behind many emerging natural and social sciences.

Considering the volume of multimedia data and difficulty in developing machine perception systems to bridge the semantic gap, it is natural that multimedia and data mining will come closer and be applied to some of the most challenging problems. And that has started to happen. Some of the

toughest challenges for data mining are posed by multimedia systems. Similarly, the potentially most rewarding applications of data mining may come from multimedia data.

As is natural and common, in the early stages of a field people explore only incremental modifications to existing approaches. And multimedia data mining is no exception. Most early tools deal with data in a single medium such as images. This is a good start, but the real challenges are in dealing with multimedia data to address problems that cannot be solved using a single medium. A major limitation of machine perception approaches, so obvious in computer vision but equally common in all other signal based systems, is their over reliance on a single medium. By using multimedia data, one can use an analysis context that is created by a data set of a medium to solve complex problems using data from other media. In a way, multimedia data mining could become a field where analysis will proceed through mutual context propagation approaches. I do hope that some young researchers will be motivated to address these rewarding areas.

This book is the very first monograph on multimedia data mining. The book presents the state-of-the-art materials in the area of multimedia data mining with three distinguishing features. First, this book brings together the literature of multimedia data mining and defines what this area is about, and puts multimedia data mining in perspective compared to other, more well-established research areas.

Second, the book includes an extensive coverage of the foundational theory of multimedia data mining with state-of-the-art materials, ranging from feature extraction and representations, to knowledge representations, to statistical learning theory and soft computing theory. Substantial effort is spent to ensure that the theory and techniques included in the book represent the state-of-the-art research in this area. Though not exhaustive, this book has a comprehensive systematic introduction to the theoretical foundations of multimedia data mining.

Third, in order to showcase to readers the potential and practical applications of the research in multimedia data mining, the book gives specific applications of multimedia data mining theory in order to solve real-world multimedia data mining problems, ranging from image search and mining, to image annotation, to video search and mining, and to audio classification.

While still in its infant stage, multimedia data mining has great momentum to further develop rapidly. It is hoped that the publication of this book shall lead and promote the further development of multimedia data mining research in academia, government, and industries, and its applications in all the sectors of our society.

Ramesh Jain
University of California at Irvine

About the Authors

Zhongfei (Mark) Zhang is an associate professor in the Computer Science Department at the State University of New York (SUNY) at Binghamton, and the director of the Multimedia Research Laboratory in the Department. He received a BS in Electronics Engineering (with Honors), an MS in Information Sciences, both from Zhejiang University, China, and a PhD in Computer Science from the University of Massachusetts at Amherst. He was on the faculty of the Computer Science and Engineering Department, and a research scientist at the Center of Excellence for Document Analysis and Recognition, both at SUNY Buffalo. His research interests include multimedia information indexing and retrieval, data mining and knowledge discovery, computer vision and image understanding, pattern recognition, and bioinformatics. He has been a principal investigator or co-principal investigator for many projects in these areas supported by the US federal government, the New York State government, as well as private industries. He holds many inventions, has served as a reviewer or a program committee member for many conferences and journals, has been a grant review panelist every year since 2000 for the federal government funding agencies (mainly NSF and NASA), New York State government funding agencies, and private funding agencies, and has served on the editorial board for several journals. He has also served as a technical consultant for a number of industrial and governmental organizations and is a recipient of several prestigious awards.

Ruofei Zhang is a computer scientist and technical manager at Yahoo! Inc. He has led the relevance R&D in Yahoo! Video Search and the contextual advertising relevance modeling and optimization group in Search & Advertising Science at Yahoo!. When he was in graduate school, he worked as a research intern at Microsoft Research Asia. His research fields are in machine learning, large scale data analysis and mining, optimization, and multimedia information retrieval. He has published over two dozen peer-reviewed academic papers in leading international journals and conferences, has written several invited papers and book chapters, has filed 10 patents on search relevance, ranking function learning, multimedia content analysis, and has served as a reviewer or a program committee member for many prestigious international journals and conferences. He is a Member of IEEE, a member of the IEEE Computer Society, and a member of ACM. He received a PhD in Computer Science with a Distinguished Dissertation Award from the State University of New York at Binghamton.

Preface

Multimedia data mining is a very interdisciplinary and multidisciplinary area. This area was developed under the two parent areas — *multimedia* and *data mining*. Since both parent areas are considered young areas with the history of around the last ten years or so, the formal development of multimedia data mining was not even established until very recently. This book is the very first monograph in the general area of multimedia data mining written in a self-contained format. This book addresses both the fundamentals and the applications of multimedia data mining. It gives a systematic introduction to the fundamental theory and concepts in this area, and at the same time, also presents specific applications that showcase the great potential and impacts for the technologies generated from the research in this area.

The authors of this book have been actively working in this area for years, and this book is the final culmination of their years of long research in this area. This book may be used as a collection of research notes for researchers in this area, a reference book for practitioners or engineers, as well as a textbook for a graduate advanced seminar in this area or any related areas. This book may also be used for an introductory course for graduate students or advanced undergraduate seniors. The references collected in this book may be used as further reading lists or references for the readers.

Due to the very interdisciplinary and multidisciplinary nature of the area of multimedia data mining, and also due to the rapid development in this area in the recent years, it is by no means meant to be exhaustive to collect complete information in this area. We have tried our best to collect the most recent developments related to the specific topics addressed in this book in the general area of multimedia data mining. For those who have already been in the area of multimedia data mining or who already know what this area is about, this book serves the purpose of a formal and systematic collection to connect all of the dots together. For those who are beginners to the area of multimedia data mining, this book serves the purpose of a formal and systematic introduction to this area.

It is not possible for us to accomplish this book without the great support from a large group of people and organizations. In particular, we would like to thank the publisher — Taylor & Francis/CRC Press for giving us the opportunity to complete this book for the readers as one of the books in the Chapman & Hall/CRC *Data Mining and Knowledge Discovery* series, with Prof. Vipin Kumar at the University of Minnesota serving as the series editor. We would like to thank this book's editor of Taylor & Francis Group, Randi Cohen, for

her enthusiastic and patient support, effort, and advice; the project editor of Taylor & Francis Group, Judith M. Simon, and the anonymous proof-reader for their meticulous effort in correcting typos and other errors of the draft of the book; and Shashi Kumar of International Typesetting and Composition for his prompt technical support in formatting the book. We would like to thank Prof. Ramesh Jain at the University of California at Irvine for the strong support to this book and kindly offering to write a foreword to this book. We would like to thank Prof. Ying Wu at Northwestern University and Prof. Chabane Djeraba at the University of Science and Technology of Lille, France, as well as another anonymous reviewer, for their painstaking effort to review the book and their valuable comments to substantially improve the quality of this book. Part of the book is derived from the original contributions made by the authors of the book as well as a group of their colleagues. We would like to specifically thank the following colleagues for their contributions: Jyh-Herng Chow, Wei Dai, Alberto del Bimbo, Christos Faloutsos, Zhen Guo, Ramesh Jain, Mingjing Li, Wei-Ying Ma, Florent Masseglia, Jia-Yu (Tim) Pan, Ramesh Sarukkai, Eric P. Xing, and HongJiang Zhang. This book project is supported in part by the National Science Foundation under grant IIS-0535162, managed by the program manager, Dr. Maria Zemankova. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Finally, we would like to thank our families for the love and support that are essential for us to complete this book.

试读结束：需要全本请在线购买： www.ertongbook.com

Contents

I Introduction **1**

1 Introduction **3**
1.1 Defining the Area 3
1.2 A Typical Architecture of a Multimedia Data Mining System 7
1.3 The Content and the Organization of This Book 8
1.4 The Audience of This Book 10
1.5 Further Readings 11

II Theory and Techniques **13**

2 Feature and Knowledge Representation for Multimedia Data **15**
2.1 Introduction 15
2.2 Basic Concepts 16
2.2.1 Digital Sampling 17
2.2.2 Media Types 18
2.3 Feature Representation 22
2.3.1 Statistical Features 23
2.3.2 Geometric Features 29
2.3.3 Meta Features 32
2.4 Knowledge Representation 32
2.4.1 Logic Representation 33
2.4.2 Semantic Networks 34
2.4.3 Frames 36
2.4.4 Constraints 38
2.4.5 Uncertainty Representation 41
2.5 Summary 44

3 Statistical Mining Theory and Techniques **45**
3.1 Introduction 45
3.2 Bayesian Learning 47
3.2.1 Bayes Theorem 47
3.2.2 Bayes Optimal Classifier 49
3.2.3 Gibbs Algorithm 50
3.2.4 Naive Bayes Classifier 50
3.2.5 Bayesian Belief Networks 52
3.3 Probabilistic Latent Semantic Analysis 56

3.3.1 Latent Semantic Analysis . . . 57
3.3.2 Probabilistic Extension to Latent Semantic Analysis . 58
3.3.3 Model Fitting with the EM Algorithm . . . 60
3.3.4 Latent Probability Space and Probabilistic Latent Semantic Analysis . . . 61
3.3.5 Model Overfitting and Tempered EM . . . 62
3.4 Latent Dirichlet Allocation for Discrete Data Analysis . . . 63
3.4.1 Latent Dirichlet Allocation . . . 64
3.4.2 Relationship to Other Latent Variable Models . . . 66
3.4.3 Inference in LDA . . . 69
3.4.4 Parameter Estimation in LDA . . . 70
3.5 Hierarchical Dirichlet Process . . . 72
3.6 Applications in Multimedia Data Mining . . . 73
3.7 Support Vector Machines . . . 74
3.8 Maximum Margin Learning for Structured Output Space . . 81
3.9 Boosting . . . 88
3.10 Multiple Instance Learning . . . 91
3.10.1 Establish the Mapping between the Word Space and the Image-VRep Space . . . 93
3.10.2 Word-to-Image Querying . . . 95
3.10.3 Image-to-Image Querying . . . 95
3.10.4 Image-to-Word Querying . . . 96
3.10.5 Multimodal Querying . . . 96
3.10.6 Scalability Analysis . . . 97
3.10.7 Adaptability Analysis . . . 97
3.11 Semi-Supervised Learning . . . 101
3.11.1 Supervised Learning . . . 104
3.11.2 Semi-Supervised Learning . . . 106
3.11.3 Semiparametric Regularized Least Squares . . . 109
3.11.4 Semiparametric Regularized Support Vector Machines 111
3.11.5 Semiparametric Regularization Algorithm . . . 113
3.11.6 Transductive Learning and Semi-Supervised Learning 113
3.11.7 Comparisons with Other Methods . . . 114
3.12 Summary . . . 115

4 Soft Computing Based Theory and Techniques 117
4.1 Introduction . . . 117
4.2 Characteristics of the Paradigms of Soft Computing . . . 118
4.3 Fuzzy Set Theory . . . 119
4.3.1 Basic Concepts and Properties of Fuzzy Sets . . . 119
4.3.2 Fuzzy Logic and Fuzzy Inference Rules . . . 123
4.3.3 Fuzzy Set Application in Multimedia Data Mining . . 124
4.4 Artificial Neural Networks . . . 125
4.4.1 Basic Architectures of Neural Networks . . . 125
4.4.2 Supervised Learning in Neural Networks . . . 131

4.4.3 Reinforcement Learning in Neural Networks 136
4.5 Genetic Algorithms 140
4.5.1 Genetic Algorithms in a Nutshell 140
4.5.2 Comparison of Conventional and Genetic Algorithms for an Extremum Search 145
4.6 Summary 150

III Multimedia Data Mining Application Examples 153

5 Image Database Modeling – Semantic Repository Training 155
5.1 Introduction 155
5.2 Background 156
5.3 Related Work 157
5.4 Image Features and Visual Dictionaries 159
5.4.1 Image Features 159
5.4.2 Visual Dictionary 160
5.5 α-Semantics Graph and Fuzzy Model for Repositories 163
5.5.1 α-Semantics Graph 163
5.5.2 Fuzzy Model for Repositories 166
5.6 Classification Based Retrieval Algorithm 168
5.7 Experiment Results 170
5.7.1 Classification Performance on a Controlled Database . 170
5.7.2 Classification Based Retrieval Results 172
5.8 Summary 180

6 Image Database Modeling – Latent Semantic Concept Discovery 181
6.1 Introduction 181
6.2 Background and Related Work 182
6.3 Region Based Image Representation 185
6.3.1 Image Segmentation 185
6.3.2 Visual Token Catalog 188
6.4 Probabilistic Hidden Semantic Model 191
6.4.1 Probabilistic Database Model 191
6.4.2 Model Fitting with EM 192
6.4.3 Estimating the Number of Concepts 194
6.5 Posterior Probability Based Image Mining and Retrieval . . 194
6.6 Approach Analysis 196
6.7 Experimental Results 199
6.8 Summary 205

7 A Multimodal Approach to Image Data Mining and Concept Discovery 209
7.1 Introduction 209
7.2 Background 210

7.3 Related Work . . . 211
7.4 Probabilistic Semantic Model . . . 213
7.4.1 Probabilistically Annotated Image Model . . . 213
7.4.2 EM Based Procedure for Model Fitting . . . 215
7.4.3 Estimating the Number of Concepts . . . 216
7.5 Model Based Image Annotation and Multimodal Image Mining and Retrieval . . . 217
7.5.1 Image Annotation and Image-to-Text Querying . . . 217
7.5.2 Text-to-Image Querying . . . 218
7.6 Experiments . . . 219
7.6.1 Dataset and Feature Sets . . . 220
7.6.2 Evaluation Metrics . . . 221
7.6.3 Results of Automatic Image Annotation . . . 221
7.6.4 Results of Single Word Text-to-Image Querying . . . 224
7.6.5 Results of Image-to-Image Querying . . . 224
7.6.6 Results of Performance Comparisons with Pure Text Indexing Methods . . . 226
7.7 Summary . . . 228

8 Concept Discovery and Mining in a Video Database 231
8.1 Introduction . . . 231
8.2 Background . . . 232
8.3 Related Work . . . 233
8.4 Video Categorization . . . 235
8.4.1 Naive Bayes Classifier . . . 237
8.4.2 Maximum Entropy Classifier . . . 238
8.4.3 Support Vector Machine Classifier . . . 240
8.4.4 Combination of Meta Data and Content Based Classifiers . . . 241
8.5 Query Categorization . . . 242
8.6 Experiments . . . 244
8.6.1 Data Sets . . . 244
8.6.2 Video Categorization Results . . . 246
8.6.3 Query Categorization Results . . . 251
8.6.4 Search Relevance Results . . . 253
8.7 Summary . . . 255

9 Concept Discovery and Mining in an Audio Database 257
9.1 Introduction . . . 257
9.2 Background and Related Work . . . 258
9.3 Feature Extraction . . . 260
9.4 Classification Method . . . 263
9.5 Experimental Results . . . 263
9.6 Summary . . . 269

References **271**

Index **291**

List of Tables

3.1 Associated conditional probabilities with the node "Alarm" in Figure 3.1. 54
3.2 Most frequently used kernel functions. 105

4.1 Comparative characteristics of the components of soft computing. Reprint from [8] ©2001 World Scientific. 119

5.1 Results of the classification tree based image classification experiments for the controlled database. Legend: A – Africa, B – Beach, C – Buildings, D – Buses, E – Dinosaurs, F – Elephants, G – Flowers, H – Horses, I – Mountains, and J – Foods. Reprint from [238] ©2004 IEEE Computer Society Press. 172
5.2 Results of the nearest-neighbor based image classification experiments for the controlled database. Legend: A – Africa, B – Beach, C – Buildings, D – Buses, E – Dinosaurs, F – Elephants, G – Flowers, H – Horses, I – Mountains, and J – Foods. Reprint from [238] ©2004 IEEE Computer Society Press. 173
5.3 The classification statistics of our method and the nearest-neighbor method. 177
5.4 The classification and retrieval precision statistics. Reprint from [244] ©2004 ACM Press and from [238] ©2004 IEEE Computer Society Press. 177

6.1 Examples of the 96 categories and their descriptions. Reprint from [243] ©2007 IEEE Signal Processing Society Press. . . . 200

7.1 Comparisons between the examples of the automatic annotations generated by the proposed prototype system and MBRM. Reprint from [246] ©2006 Springer-Verlag Press and from [245] ©2005 IEEE Computer Society Press. 222
7.2 Performance comparison on the task of automatic image annotation on the test set. Reprint from [246] ©2006 Springer-Verlag Press and from [245] ©2005 IEEE Computer Society Press. 224

8.1 Matrix representation of users' query log. Reprint from [239] ©2006 ACM Press. 243

试读结束：需要全本请在线购买： www.ertongbook.com