

世界著名计算机教材精选

多媒体信息 检索与管理

David Dagan Feng
Wan-Chi Siu
Hong-Jiang Zhang

著

刘晓冬 译



MULTIMEDIA INFORMATION RETRIEVAL AND MANAGEMENT

Technological Fundamentals and Applications



清华大学出版社

世界著名计算机教材精选

Multimedia Information Retrieval and Management
Technological Fundamentals and Applications

多媒体信息检索与管理

David Dagan Feng
Wan-Chi Siu 著
Hong-Jiang Zhang
刘晓冬 译

清华大学出版社
北 京

English reprint edition copyright © 2009 by Springer-Verlag and TSINGHUA UNIVERSITY PRESS.

Original English language title from Proprietor's edition of the Work.

Original English language title: Multimedia Information Retrieval and Management: Technological Fundamentals and Applications by David Dagan Feng, Wan-Chi Siu, Hong-Jiang Zhang, Copyright © 2008

All Rights Reserved.

This edition has been authorized by Springer-Verlag (Berlin/Heidelberg/New York) for sale in the People's Republic of China only and not for export therefrom.

本书翻译版由 Springer-Verlag 授权给清华大学出版社出版发行。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

多媒体信息检索与管理/冯大淦,萧允治,张宏江著;刘晓冬译. —北京:清华大学出版社, 2009.11

(世界著名计算机教材精选)

书名原文: Multimedia Information Retrieval and Management: Technological Fundamentals and Applications

ISBN 978-7-302-21161-7

I. 多… II. ①冯… ②萧… ③张… ④刘… III. 情报检索—高等学校—教材
IV. G252.7

中国版本图书馆 CIP 数据核字(2009)第 182770 号

责任编辑:龙啟铭

责任校对:徐俊伟

责任印制:何 芊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:北京市世界知识印刷厂

装 订 者:三河市李旗庄少明装订厂

经 销:全国新华书店

开 本:185×260

印 张:26.25

字 数:620 千字

版 次:2009 年 11 月第 1 版

印 次:2009 年 11 月第 1 次印刷

印 数:1~3000

定 价:59.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。
联系电话:010-62770177 转 3103 产品编号:027766-01

David Dagan Feng (冯大淦) 于 1988 年在美国加州大学洛杉矶分校获得计算机科学博士学位。在美国短暂地任助理教授一段时间后, 加入了澳大利亚悉尼大学并先后任讲师、高级讲师、副教授和教授。他是悉尼大学计算机科学系/信息技术学院的前系主任。他现任悉尼大学理学院的副院长和悉尼大学生物医学和多媒体信息 (BMIT) 研究组的主任, 以及香港理工大学电子和信息工程系多媒体信号处理研究中心的教授和副主任。他已经发表了 200 多篇学术论文, 开创了几个新的研究方向, 在他的领域作出了许多重大的贡献。因为在生物医学工程方面的杰出成就就获得了 Crump 奖。然而更重要的是, 他的很多研究成果已经被转化为世界性的现实生活问题的解决方案, 并已对人们的生活质量做出了巨大的改善。他现在是 IEEE Transactions on Information Technology in Biomedicine 的专题领域编辑, International Journal of Image and Graphics 的顾问, IFAC-TC-BIOMED 的主席, 以及 IFAC 2003 研讨会的国际程序和国家组织委员会主席。他是 ACS、HKIE、IEE 和 IEEE 的院士。

Wan-Chi Siu (萧允治) 于 1984 年在英国伦敦帝国理工和医学院获得博士学位。1975—1980 年在香港中文大学工作。1980 年加入香港理工大学任讲师, 自 1992 年开始成为讲座教授。他还承担了管理工作, 1992—1994 年, 他任工程学院副院长。1994—2000 年任电子和信息工程系主任, 2000—2002 年任工程学院院长。Siu 教授自 1998 年 9 月以来一直担任多媒体信号处理中心主任。已发表研究论文 200 多篇, 他的研究兴趣包括数字信号处理、快速计算算法、变换、图像和视频编码, 以及模式识别和神经网络的计算方面。Siu 教授是信号、图像和视频技术的 VLSI 信号处理系统学报和应用信号处理的 EURASIP 学报的编委会成员。他是 1998 年 5 月出版的 IEEE Transactions on Circuits and System Pt. II 专刊的客座编辑, 以及在 1995—1997 年之间出版的 IEEE Transactions on Circuits and System Pt. II 专刊的副编辑。Siu 教授担任过许多国际会议的大会主席或技术程序主席。特别地, 他是 IEEE 电路和系统国际研讨会 (ISCAS'97) 的技术程序主席, 以及在香港举行的 2003 IEEE 声学、语音和信号处理国际会议 (ICASSP'2003) 的大会主席。1991—1995 年期间, Siu 教授是香港政府科研经费委员会 (RGC) 物

理科学和工程专家组的成员。1994 年他担任第一个工程和信息技术专家组席, 评估了香港全部大学中的 19 个成本中心(系)的研究质量。

Hong-Jiang Zhang (张宏江) 于 1991 年在丹麦 Lyngby 的丹麦工程技术大学获得博士学位。他在新加坡国立大学系统科学研究所和美国惠普实验室担任研究管理者。Zhang 博士于 1999 年加入微软亚洲研究院(MRA), 担任高级研究员和 MRA 院长助理。Zhang 博士在多媒体研究领域以他在视频和图像内容分析、表达、检索和浏览方面的开创性工作而闻名, 已经发表了 3 本著作, 200 篇被引用的论文和书的章节, 以及大量的在多媒体处理和基于内容的媒体检索领域的专业学报的专辑。他有 30 多项美国专利和专利申请。Zhang 博士是《多媒体系统图像和视频处理》的作者之一, 这是第一本介绍基于内容的图像和视频检索研究的著作。该书自 1995 年在 Kluwer 出版社出版以来, 已经成为经典参考教材, 并且包含在其中的研究已经成为几个初创公司的技术基础。Zhang 博士是 5 个专业学报和 12 个国际会议的编委会成员, 包括 ACM 1999 多媒体会议程序委员会共同主席。

序言

多媒体信息技术为广泛的应用领域提供了全面的和直观的信息，对现代生活有强烈的影响并已改变了我们的学习和思考方式。近 20 年来，在 Internet 和无线通信上对数字多媒体（包括音频、视频、图像和图形）的使用有了一个爆炸性的增长。随着对数字多媒体使用的增加，有效的数据存储和管理变得越来越重要。在使用大量数据的领域（例如，音频、视频、图像和数字图书馆；地理和医学图像数据库等），我们需要在满足对精确的数据表达经常相互冲突的需求的同时，最小化存储的数据量。此外，数据需要被管理从而促进高效的搜索、浏览和协同工作。这个领域在近年来已经是一个非常活跃的研究领域。本书将为读者提供那些直接影响我们的工业、经济和社会生活的多媒体信息检索和管理领域的前沿技术的最新的和综合的描绘。

本书分成两个主要的部分：覆盖该领域核心理论的技术基础部分和描述该技术广泛的实际使用的应用部分。全书包括：

- 对于数字音频、音乐、图像和视频基于内容检索的一套完整的技术；
- 多媒体低级特征提取和高级语义描述，以及最新的 MPEG-7 标准；
- 多媒体认证和水印；
- 广泛的实际应用（例如，数字图书馆、医学图像和生物特征识别等）。

本书适合于用作高年级大学生和研究生的教材，也可以作为多媒体研究者、开发者和从业者的参考手册。

本书的贡献者都是在他们的领域中的杰出的研究者，包括：哥伦比亚大学的 Shih-Fu Chang 教授，他是 IEEE 2002 年的杰出演讲者；南加州大学的 H. K. Huang 教授，他是“图片存档和通信系统 (PACS)”的作者和 PACS 系统的开拓者，以及 Edward A. Fox 教授，他是弗吉尼亚理工大学因特网技术创新中心的主任和 ACM 计算教育资源学报的共同主编。我们要感谢这些贡献者的出色的配合，使得我们能够满足很紧的撰写截止时间，也确保了本书所包含的信息是最新的。我们还要感谢 Z. Y. Wang 先生和 W. Ho 女士在编辑这本书时给予的专业协助，并感谢 Springer-Verlag 出版社在整个过程中的支持和效率。

译者序

由德国 Springer-Verlag 出版社出版的 David Dagan Feng、Wan-Chi Siu 和 Hong-Jiang Zhang 等编著的本书，是一本关于多媒体信息检索和管理领域的经典著作，倾注了该领域世界众多知名学者的研究成果和贡献，在国外许多重点大学作为计算机专业本科生和研究生的教材。本书填补了我国在该研究领域译著的空白。

这本书在内容安排上，从理论基础和实际应用两个方面对当前流行的多媒体信息检索和管理领域的技术和系统进行了完整和详细的介绍。全书共分两大部分：第 1 部分是技术基础篇，包括第 1 章基于内容的图像检索基础，第 2 章基于内容的视频分析、检索和浏览，第 3 章基于内容的图像检索的相关反馈，第 4 章结构和语义层次的视频分析和摘要，第 5 章数字音频和音乐的基于内容的检索，第 6 章 MPEG-7 多媒体内容描述标准，第 7 章多媒体认证和水印技术，第 8 章索引和检索高维视觉特征，第 9 章多媒体数据库的语义数据建模机制和第 10 章可视化数据库系统中的基于特征的检索。第 2 部分是应用篇，包括第 11 章数字图书馆，第 12 章数字图书馆的可扩展存储，第 13 章数字图书馆的对象知识库，第 14 章万维网上的信息发现，第 15 章协同多媒体信息系统，第 16 章图片存档及通信系统，图像管理和影像信息学，第 17 章医学数据基于内容的检索，第 18 章使用掌纹图像的生物特征检索，第 19 章在人脸数据库中查找人脸，第 20 章活体植物识别的数据管理和第 21 章广播视频的快速启动与交互式检索。

与传统的多媒体教材不同，本书的特点是介绍在 Internet、无线通信以及使用大量数据的领域（如数字图书馆、地理和医学图像数据库等）中对数字多媒体（包括音频、视频、图像和图形）数据的有效存储和管理以促进高效的检索、浏览和协同工作。这本书融合了多媒体信息检索和管理领域的众多前沿技术，并在每章结尾列出了大量的参考文献以供读者根据自己的研究兴趣作进一步更深入的研究。

本书适用读者很广泛。对于想学习和了解计算机多媒体信息检索和管理的本科生、研究生，本书可作为他们的教材或教学参考书。对于那些从事计算机多媒体信息检索和管理工程领域的研究开发人员、技术服务人员、用户等，本书也是一本很好的参考书。

目前国内在计算机多媒体信息检索和管理领域的教材尚少。在这种背景下，由清华大学出版社和德国 Springer-Verlag 出版社合作引进这本在国

外已被广泛使用、内容先进的多媒体信息检索和管理著作是一件有益的事情。广大读者可通过本书系统地学习多媒体信息检索和管理领域的知识。相信随着这本著作的翻译出版，必将为我国多媒体信息检索和管理领域的相关工作人员提供一本很好的参考书。

感谢王瑞、傅崇刚、刘芳、王鹏、韦奕多在本书的翻译和校对过程中给予的帮助和支持。也感谢我的妻子和女儿在我翻译过程中给我的鼓励和支持。由于本书涉及的学科范围广、内容众多，翻译难度较大，加之译者水平有限，书中难免有疏漏或错误，欢迎广大读者批评指正。

刘晓冬

2009年5月于清华大学

目录

第1部分 技术基础篇

第1章 基于内容的图像检索基础	3
1.1 引言	3
1.2 图像内容描述符	4
1.2.1 颜色	5
1.2.2 纹理	8
1.2.3 形状	12
1.2.4 空间信息	15
1.3 相似度测量和索引方案	15
1.3.1 相似度/距离测量	15
1.3.2 索引方案	17
1.4 用户交互	18
1.4.1 查询说明	18
1.4.2 相关反馈	18
1.5 性能评价	19
1.6 结论	20
参考文献	21
第2章 基于内容的视频分析、检索和浏览	26
2.1 引言	26
2.2 视频结构分析	27
2.2.1 镜头边界检测	28
2.2.2 场景分组	30
2.2.3 情节分割	30
2.3 视频内容表达	31
2.3.1 基于关键帧的表达	32
2.3.2 基于镜头的表达	32
2.4 视频提取和视频摘要	39
2.4.1 关键帧抽取	39
2.4.2 视频高亮显示	41

2.5 基于内容的视频检索和浏览.....	42
2.5.1 相似性度量	42
2.5.2 镜头聚类	44
2.5.3 浏览方案	44
2.6 结论	46
致谢	47
参考文献	47
 第3章 基于内容的图像检索的相关反馈	 51
3.1 引言	51
3.2 相关反馈算法	52
3.2.1 经典算法	52
3.2.2 作为学习问题的相关反馈	53
3.2.3 带记忆的相关反馈	56
3.3 一个集成相关反馈框架	57
3.3.1 语义网络	58
3.3.2 集成和跨模态查询和检索	59
3.3.3 概率传播方案	60
3.3.4 实验结果	61
3.4 结论	62
致谢	63
参考文献	63
 第4章 结构和语义层次的视频分析和摘要	 66
4.1 引言	66
4.2 场景分析	67
4.2.1 计算场景定义	67
4.2.2 方法	69
4.3 事件分析	71
4.3.1 方法	71
4.4 视频摘要	73
4.4.1 图像故事板	73
4.4.2 缩略视频	74
4.4.3 “最优”摘要存在吗	77
4.5 结论	78
致谢	79
参考文献	79

第 5 章 数字音频和音乐的基于内容的检索	83
5.1 引言	83
5.2 音频特征抽取	84
5.2.1 声学特征	84
5.2.2 主观/语义特征	85
5.3 一般音频分类和检索	86
5.3.1 基于内容的音频分割	86
5.3.2 基于内容的音频检索	87
5.4 音乐内容分析	88
5.4.1 音乐分割	88
5.4.2 音乐类型分类	89
5.4.3 音乐摘要	89
5.4.4 乐器鉴定	90
5.4.5 音乐标注	91
5.5 基于内容的音乐检索	92
5.6 案例研究	93
5.6.1 基于内容的音频检索	93
5.6.2 基于内容的音乐哼唱检索	97
5.7 结论	100
5.7.1 从录音中提取旋律	100
5.7.2 检索精度	101
5.7.3 基于 Web 的搜索引擎	101
致谢	101
参考文献	101
第 6 章 MPEG-7 多媒体内容描述标准	105
6.1 引言	105
6.2 背景和应用	106
6.2.1 MPEG 背景	106
6.2.2 MPEG-7 应用	107
6.3 MPEG-7 结构	109
6.3.1 MPEG-7 组成部分	109
6.3.2 MPEG-7 基本结构	109
6.3.3 MPEG-7 可扩展性	110
6.4 MPEG-7 概念模型	111
6.5 MPEG-7 描述定义语言	112



6.6	MPEG-7 多媒体描述方案	113
6.6.1	基本元素	114
6.6.2	内容描述工具	114
6.6.3	内容组织、导航和访问、用户交互	114
6.6.4	分类方案	115
6.6.5	例子	115
6.7	MPEG-7 音频描述工具	119
6.7.1	低层次音频工具	119
6.7.2	高层次音频工具	120
6.7.3	例子	120
6.8	MPEG-7 视觉描述工具	122
6.8.1	颜色	122
6.8.2	纹理	122
6.8.3	形状	123
6.8.4	运动	123
6.8.5	定位	123
6.8.6	人脸	123
6.8.7	例子	123
6.9	MPEG-7 的未来	125
	参考文献	125
第 7 章	多媒体认证和水印技术	129
7.1	引言	129
7.2	多媒体认证范例	131
7.2.1	需求	131
7.2.2	对多媒体认证系统的攻击	132
7.3	多媒体认证方案	133
7.3.1	使用脆弱水印的硬多媒体认证	133
7.3.2	使用半脆弱水印的软多媒体认证	137
7.3.3	基于内容的认证	140
7.4	结论	147
	参考文献	147
第 8 章	索引和检索高维视觉特征	151
8.1	引言	151
8.2	R 树家族	154
8.3	CSS+树	155
8.3.1	CSS+树的一般结构	155

8.3.2	CSS+树的批量加载	156
8.3.3	CSS+树的动态行为	158
8.3.4	同其他层次化结构的比较	160
8.4	处理不同的距离度量	161
8.4.1	用加权欧氏距离进行 k-nn 搜索的方法	162
8.4.2	到高维空间的扩展	163
8.4.3	实验结果和讨论	163
8.5	降维	164
8.5.1	使用混合联想神经网络降维	165
8.5.2	降维结果	166
8.6	结论	168
	参考文献	169
第 9 章	多媒体数据库的语义数据建模机制	172
9.1	引言	172
9.2	语义和多媒体数据建模技术	174
9.2.1	多媒体数据库技术	174
9.2.2	面向对象的视图机制	174
9.2.3	动态对象模型	175
9.3	<i>MediaView</i> : 一种语义建模机制	176
9.3.1	概貌	176
9.3.2	<i>MediaView</i> 的基本原理	177
9.3.3	视图算子	180
9.3.4	视图推导机制	181
9.4	<i>MediaView</i> 的应用	184
9.4.1	多媒体查询	184
9.4.2	“媒体地图”中的导航	186
9.4.3	存储和索引的优化	187
9.5	结论	187
	参考文献	188
第 10 章	可视化数据库系统中基于特征的检索	190
10.1	引言	190
10.2	可视化数据库系统设计	191
10.3	元数据层次	193
10.3.1	特征索引	194
10.3.2	聚类和模板	196
10.3.3	异类特征的聚类	196

10.4 异类特征的查询.....	199
10.4.1 使用异类特征排序图像	200
10.4.2 合并异类特征.....	202
10.4.3 相关反馈的集成	202
10.5 结论.....	203
致谢.....	204
参考文献.....	204

第2部分 应用篇

第11章 数字图书馆.....	209
11.1 数字图书馆定义.....	209
11.2 数字图书馆创新工程（第一阶段）	211
11.3 DLI 项目	212
11.3.1 卡内基梅隆大学“视频的全内容搜索和检索”	212
11.3.2 伊利诺斯大学乌尔班纳-香巴尼分校“科学文献的联合知识库”	215
11.3.3 斯坦福大学“异类服务中的互操作机制”	217
11.4 DLI 二期.....	219
11.4.1 以人为中心的研究.....	220
11.4.2 基于内容和馆藏的研究.....	220
11.4.3 以系统为中心的研究.....	221
11.4.4 建立测试平台	221
11.4.5 总结.....	222
11.5 结论.....	222
参考文献.....	222
第12章 数字图书馆的可扩展存储.....	225
12.1 引言.....	225
12.2 动机.....	226
12.3 主题领域.....	226
12.4 本地文件系统.....	227
12.5 工作负载研究.....	227
12.6 本地文件系统性能问题.....	230
12.6.1 高速缓存.....	230
12.6.2 聚集和分裂.....	231
12.6.3 块尺寸和分配.....	231
12.6.4 日志结构的方法	232

12.7 磁盘阵列.....	232
12.8 网络化存储.....	236
12.8.1 智能磁盘.....	237
12.8.2 并行文件系统.....	238
12.8.3 分布式文件系统.....	239
12.9 结论.....	241
参考文献.....	241
第 13 章 数字图书馆的对象知识库.....	247
13.1 引言.....	247
13.2 数字对象与文件.....	248
13.3 命名和定位.....	249
13.3.1 唯一性与定位依赖性.....	249
13.3.2 统一资源名称.....	250
13.3.3 可扩展的对象定位.....	251
13.4 可靠性冗余编码.....	252
13.5 元数据.....	253
13.5.1 桶.....	254
13.5.2 条款和条件.....	254
13.6 数字对象知识库.....	255
13.6.1 KAHN-WILENSKY 及其扩展.....	256
13.6.2 其他的知识库方法.....	257
13.7 结论.....	258
参考文献.....	259
第 14 章 万维网上的信息发现.....	263
14.1 引言.....	263
14.2 Web 结构挖掘.....	267
14.2.1 权威和枢纽.....	267
14.2.2 锚点.....	268
14.2.3 网页排名.....	271
14.3 Web 日志挖掘.....	272
14.3.1 用户模型.....	272
14.3.2 因特网 GLOSS.....	273
14.4 结论.....	276
参考文献.....	276

第 15 章 协同多媒体信息系统	278
15.1 引言	278
15.1.1 协同多媒体信息系统定义	279
15.1.2 多媒体信息系统的分类	279
15.2 协同多媒体信息系统回顾	280
15.2.1 多媒体会议	280
15.2.2 集成工具集	283
15.3 体系结构实例	285
15.3.1 MARRATECH	285
15.3.2 多媒体交互式远程学习系统	286
15.3.3 会议纪要系统	287
15.3.4 TEAMSMART	287
15.4 管理问题	289
15.5 结论	291
参考文献	291
第 16 章 图片存档及通信系统、图像管理和影像信息学	297
16.1 图片存档及通信系统 (PACS) 基础	297
16.1.1 PACS 组件	297
16.1.2 医学图像	299
16.1.3 术语和标准	300
16.1.4 IHE 和 PACS 数据流	301
16.1.5 HIS/CMS 和电子病历	301
16.2 图像数据通信和管理	302
16.2.1 图像通信	302
16.2.2 大尺度图像数据管理	305
16.3 影像信息学	307
16.3.1 医学影像信息学基础设施	307
16.3.2 NCM—神经外科病人数据的管理	309
16.4 总结	311
16.5 结论	311
致谢	312
参考文献	312
第 17 章 医学数据基于内容的检索	313
17.1 引言	313
17.2 一维医学数据基于内容的检索	314

17.3 二维医学数据基于内容的检索.....	315
17.4 三维医学数据基于内容的检索.....	316
17.5 四维医学数据基于内容的检索.....	318
17.5.1 示踪剂动力学特征提取和数据压缩	319
17.5.2 四维功能图像检索系统	321
17.5.3 案例研究.....	323
17.6 结论.....	324
致谢.....	325
参考文献.....	325
第 18 章 使用掌纹图像的生物特征检索	329
18.1 引言.....	329
18.2 动态特征选择.....	335
18.3 分式判别函数.....	336
18.4 分层掌纹匹配.....	337
18.5 实验结果.....	340
18.6 结论.....	341
致谢.....	342
参考文献.....	342
第 19 章 在人脸数据库中查找人脸	344
19.1 人脸识别: 综述和挑战.....	344
19.1.1 自动人脸识别的应用	344
19.1.2 人脸识别系统的结构	345
19.2 特征脸.....	346
19.3 复杂背景中的人脸检测.....	347
19.3.1 颜色空间.....	347
19.3.2 类似人脸区域的分割	348
19.4 面部特征检测和提取.....	352
19.4.1 提取人脸轮廓.....	353
19.4.2 变形模板.....	355
19.5 从人脸数据库中搜索人脸.....	358
19.5.1 基于特征脸的人脸识别	358
19.5.2 使用豪斯多夫距离度量的人脸识别	359
19.6 结论.....	363
参考文献.....	364