信息科学与工程系列专著

跨媒体信息技术导论

杨毅　王胜开　陈国顺　徐为群　马欣　编著

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

信息科学与工程系列专著

跨媒体信息技术导论

杨毅　王胜开　陈国顺　徐为群　马欣　编著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书首先介绍文本、语音、图像、视频等传统多媒体信息处理的基本概念、基本理论、基本模型，而后论述跨媒体信息处理这一新技术领域的国内外最新发展状况与研究成果，分析跨媒体信息表示、检索和处理的基本理论、数学模型和关键技术等，描述跨媒体信息系统的基本结构，并结合典型系统，对跨媒体信息处理技术的应用与发展做了介绍。

本书可供信息技术、模式识别、信号处理、多媒体与跨媒体信息处理等领域的工程技术人员、科研管理人员以及相关专业的大专院校学生、研究生参考。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

跨媒体信息技术导论 / 杨毅等编著. —北京：电子工业出版社，2012.11
（信息科学与工程系列专著）
ISBN 978-7-121-18670-7

Ⅰ. ①跨… Ⅱ. ①杨… Ⅲ. ①传播媒介－信息技术 Ⅳ. ①G20

中国版本图书馆 CIP 数据核字（2012）第 238129 号

责任编辑：田宏峰
印　　刷：三河市鑫金马印装有限公司
装　　订：三河市鑫金马印装有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编　100036
开　　本：787×980　1/16　印张：11.5　字数：257 千字
印　　次：2012 年 11 月第 1 次印刷
印　　数：3 000 册　定价：39.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

前　言

信息时代，随着计算机、互联网和数字媒体的不断发展，以多种媒体形式呈现的信息急剧增加，用户对多媒体信息的应用越来越普遍。但在面对浩瀚的信息海洋时，人们也正面临日益复杂的信息筛选与信息整合困境，针对多形式媒体信息的表示、检索、提取和处理问题越来越引起人们的关注。海量多形式媒体信息的处理涉及信息学、心理学、信号处理、模式识别、信息论、控制论等众多学科和技术领域，是目前的一个跨学科研究热点，正逐步发展成为一个专业的研究领域，即跨媒体信息处理。

跨媒体指的是多维、高阶和海量的文本、语音、图像和视频等信息，这些信息蕴涵广泛而复杂的交叉关联性。跨媒体信息处理涉及有效的文本、语音、图像与视频数据挖掘、海量数据存储、搜索引擎、数据实时分析、跨媒体统一表示与建模、跨媒体信息智能处理与全局融合、跨媒体信息挖掘与知识获取的模型和方法、跨媒体信息存取与知识表达的索引方法、跨媒体信息实时处理与验证等众多理论和技术问题。

传统的基于多媒体（如文本、语音、图像、视频等）的信息处理技术，作为相对独立的学科方向，已逐渐成为相对独立的研究领域，虽然有相近的理论和算法作为研究手段和技术，但到目前为止仍缺乏对不同种类信息之间的关联性理解、表示、分析和处理手段，现有技术在跨媒体信息检索、管理和预测等方面已经不能满足需要，必须在现有技术的基础上，探索和寻求新的跨媒体信息处理方法，以便更好地实现多种形式媒体信息的处理和融合，提取更加丰富的内容，以更加多彩、自然的形式呈现在人们面前。

本书首先介绍文本、语音、图像等形式的多媒体信息处理的基本概念、基本理论、基本方法，使读者在对这些领域发展概貌有比较全面的了解后，通过对跨媒体信息处理模型和方法等的系统介绍和深入分析，使读者对跨媒体信息处理这一新的技术领域有比较全面的认识。

全书主要包括三大部分内容：第一部分为绪论，介绍跨媒体信息处理技术的基本概念，以及国内外目前研究状况；第二部分为多媒体信息处理基础知识，对文本、语音、图像、视频等形式媒体信息处理的基本概念、基本理论、基本模型、国内外最新研究成果与技术进展等进行描述和分析，使读者对多媒体信息处理有一个全面了解；第三部分为跨媒体信息处理，主要内容包括跨媒体信息表示、检索和处理的理论、模型、技术与方法，以及跨媒体信息系统的基本组成结构、基本特点等，并通过对典型应用案例的介绍，使读者对跨媒体信息系统的发展与应用情况有更加深刻的了解。

跨媒体信息处理涉及众多新理论、新技术、新方法，目前国内外的相关参考资料和专业文献较少，许多理论和技术问题尚处于探讨与摸索阶段。因此，全书力求深入浅出、图文并茂，理论模型叙述与典型应用案例分析相结合，使不具备相关技术基础的读者也能够快速掌

握相关知识，尤其是信息技术、模式识别、信号处理、多媒体与跨媒体信息处理等领域的工程技术人员、科研管理人员以及相关专业的大专院校学生、研究生参考，使更多的读者认识和研究这一新领域、新问题，共同推动跨媒体信息处理技术的发展。

在本书编写过程中，清华大学刘润生教授、清华大学宋健教授、中国科学院声学研究所颜永红教授、北京科技大学余达太教授等学者和专家，给予了大力指导，提供了大量资料和支持，在此表示衷心感谢！

因作者水平和经验有限，书中不当之处在所难免，敬请读者指正。

作　者

2012 年 10 月

试读结束：需要全本请在线购买：　www.ertongbook.com

目　录

第1章 绪 论

1.0 引言

随着多媒体技术的发展，计算机可以存储、分析和理解的多媒体数据不断增多，从单一的文本发展到图像、音频、视频、3D 模型等半结构化和无结构化的数据。“跨媒体”概念的提出正是基于多媒体技术的不断发展，它将更加符合人脑对视觉、听觉等不同感官信息的综合处理模式，使计算机能够更好地模拟人脑处理、管理和应用不同类型的媒体数据。

本章从人类认知基本理论入手，通过对多媒体概念的介绍，引出跨媒体信息的定义，描述跨媒体信息的表示与检索，说明跨媒体信息的应用。

1.1 人类认知基本理论

人类通过视觉、听觉、触觉等不同感官形成对事物的感知，本质上，人脑所处理的信息本身就具有跨媒体特性，“McGurk 现象”和近期神经系统科学进行的研究从不同角度揭示了人脑认知的跨媒体特性。1976 年，McGurk 等人验证了人类对外界信息的认知是基于不同感官信息（如听觉和视觉等）而形成的整体性理解，任何感官信息的缺乏或不准确将导致大脑对外界信息的理解产生偏差，这个现象被称为“McGurk 现象”[7]。McGurk 现象揭示了大脑在进行感知时，不同感官被无意识和自动地结合到了一起进行处理。更为重要的是，后续神经系统科学研究也揭示，在大脑皮层的颖上沟和脑顶内沟等部位，不同感官信息的处理神经相互交融，人脑的生理组织结构决定了其对外界的认知过程是通过跨越多种感官信息的融合处理来实现的[8]。

另外，从人工智能研究的角度来看，1976 年 Newell 和 Simon 提出了物理符号系统假设，认为物理符号系统是表现智能行为的必要和充分条件，任何信息加工系统都可以看成一个具体的物理系统，如人的神经系统、计算机的构造系统等。之后以 McCorthy 和 Nilsson 等为代表，主张任何事物都可以用统一的逻辑框架来表示，即可以用形式化的方法来描述客观世界。20 世纪 70 年代后期提出的知识系统，作为人工智能学科最重要的工业化和商业化产物，辅助人们进行问题求解，如产品质量的评价、辅助医疗诊断、金融决策支持等。传统的人工智

能研究的目标是让机器模仿人，认为人脑的思维活动可以通过一些公式和规则来定义，希望通过把人类的思维方式翻译成程序语言输入机器，使机器有朝一日能产生像人类一样的思维能力。然而，人脑得到的信息中可以符号化的只占很小一部分，85%以上是符号以外的形象数据，如一幅花红柳绿的风景图、一段余音绕梁的音乐等。传统的人工智能研究面对多媒体的信息环境，不能自如地模拟人脑的智能活动。跨媒体思想对于人工智能研究的重要意义正体现在着眼于对85%以上的非符号信息的综合理解和有效利用，以使计算机可更好地模拟人类感知。

跨媒体是一个比较广义的概念，主要涉及以下研究范畴。

1）跨媒体检索

用户向计算机提交一种类型的多媒体对象作为查询例子，系统可以自动找到其他不同类型、在语义上相似的多媒体对象。虽然不同类型的多媒体对象之间没有直接的可比性，如一幅山水画和一段描述小河流水声的音频在底层内容特征上彼此异构，但却可以用机器学习、统计分析等方法学习两者在统计意义上潜在的相关性，并以此为依据进行跨媒体检索。

2）跨媒体推理

推理是从一种命题合理演绎到另一种命题，跨媒体推理就是从一种类型的多媒体数据经过问题求解转向另一种类型的多媒体数据。例如，OCR（Optical Character Recognition）技术是从图像到文本的推理、基于内容的图像检索是从图像到图像的推理、视频动画技术是从视频数据到动画序列的演绎。跨媒体推理囊括了对这些不同类型的多媒体数据之间的转换研究。

3）跨媒体存储

现有的处理海量数据的检索技术主要针对的是文本信息，如谷歌和百度等搜索引擎，针对多媒体检索的研究工作其出发点并不是针对跨媒体海量数据。跨媒体存储研究高效压缩、索引和分片等方法，以及对用户行为的个性化索引等技术，用于提高海量环境下的跨媒体检索效率，更好地支持上层应用。

上述三个方面，从底层数据存储到上层应用技术，从不同方面描述了跨媒体思想对多媒体研究领域的技术涵盖和突破性要求，是一个整体性的研究框架设计和考虑。要实现上述研究思路，需要在海量数据库、多媒体索引、并行计算、机器学习和统计分析、计算机视觉、计算机听觉以及信息检索等领域取得突破性的研究进展。

1.2 跨媒体信息基本概念

1.2.1 多媒体基本概念

在人类社会发展中，信息的表现形式是多种多样的，通常将这些表现形式称为媒体（Medium）。使用计算机记录和传播的信息媒体都有一个共同特点，即信息的最小单元是比特

试读结束：需要全本请在线购买： www.ertongbook.com

(bit)，任何信息在计算机中存储和传播时都可以分解为一系列“0”或“1”的排列组合。通常将通过计算机存储、处理和传播的信息媒体称为数字媒体（Digital Media）[5]。

多媒体综合了计算机、图形学、图像处理、影视技术、音乐、美术、教育学、心理学、人工智能、信息学与电子技术学等众多学科与技术，它集文字、图形、图像、声音、视频影像和动画等多种信息于一体，能充分调动人的视觉和听觉处理功能。

国际电信联盟（ITU）将媒体细分为如下五种类型。

1）感觉媒体（Perception Medium）

感觉媒体是指直接作用于人的感觉器官，使人产生直接感觉的媒体。例如，引起听觉反应的声音和引起视觉反应的图像等。

2）表示媒体（Representation Medium）

表示媒体是指传输感觉媒体的中介媒体，即用于数据交换的编码。例如，图像编码(JPEG、MPEG)、文本编码（ASCII、GB2312）和声音编码等。

3）表现媒体（Presentation Medium）

表现媒体是指进行信息输入和输出的媒体。例如，键盘、鼠标、扫描仪、话筒和摄像机等都是输入媒体，显示器、打印机和喇叭等都是输出媒体。

4）存储媒体（Storage Medium）

存储媒体是指用于存储表示媒体的物理介质。例如，硬盘、光盘等。

5）传输媒体（Transmission Medium）

传输媒体是指用于传输表示媒体的物理介质。

多媒体计算机中所说的媒体，通常是指信息的表现形式（即传播形式），如文字、声音、图像和动画等。也就是说，计算机不仅能够处理文字、数值之类的信息，而且能够处理声音、图形、电视图像等多种不同形式的信息。

多媒体的概念常用来兼指多媒体信息和多媒体技术。所谓多媒体信息，是指集数据、文字、图形、图像和声音等为一体的综合媒体信息；所谓多媒体技术，是指将计算机技术与通信传播技术融为一体，综合处理、传输和存储多媒体信息的数字技术，它提供了良好的人机交互功能和可编程环境，极大地拓展了计算机应用领域，改变了人们的工作、学习、生活方式，并对大众传播媒体产生了巨大的影响。

1.2.2 基于内容的多媒体检索

信息检索的定义可以追溯到20世纪40年代，它是指信息按一定的方式组织起来，并根据信息用户的需要找出有关信息的过程和技术。狭义的信息检索就是指信息检索过程的后半

部分，即从信息集合中找出所需信息的过程，也就是我们常说的信息查询，其中至少涉及以下三点[4]：

（1）用户对要找寻信息的内容进行高度抽象概括，形成语义描述；

（2）使用一个相似度量函数，从信息仓库中得到与用户请求相似的信息集合，并将它们反馈给用户；

（3）用何种系统和何种技术自动实现上面两个目标。

当初提出信息检索概念时，人们仅仅关注纯文本的检索操作，现已有了很大的发展，如今网上有各种搜索引擎，为人们从网上获取有用的信息提供了极大的方便。然而随着多媒体的出现，信息检索已经不能只考虑纯文本的检索了，还需要考虑多种信息载体。

随着计算机的发展，人们对多媒体的需求也越来越多，全世界多媒体数据的产生速度也越来越快，如美国航空航天局的数字行星项目每天要产生 1 000 GB 的新数据。可以说每天都会不断更新庞大的数字多媒体资源，在这种情况下，纯文本资源的使用将越来越少。然而，到目前为止，人们对于多媒体还缺乏非常有效的检索手段，庞大的数字多媒体资源还显得有些杂乱无章，难以达到资源充分共享的目的。

人们对于文本信息的检索已经有了深入的研究，建立了多种匹配算法模型，主要有布尔模型、聚类模型、向量模型和概率模型。不同于文本信息，多媒体语义内容是通过多种媒质（如视频图像、音频和文字等）共同表达和补充的，因此，对多媒体信息分析就要对蕴涵在多媒体数据流内的所有媒质特征进行分析，这些媒质包括视频流中的图像帧、音频信号流、从视频图像中提取的字幕、由音频信号转录得到的语音和三维模拟物体等信息。在对这些媒质提取特征后，就可以使用这些提取的特征来表征原有媒质，进而将连续的多媒体数据流分割成有语义信息的单位（如镜头和场景、语音与音乐等），最后将这些语义单位识别分类成先前定义的模板类型，为它们建立索引，以便以后的检索与浏览。

基于内容的多媒体分析检索是指对多媒体数据（如视频、音频流等）所蕴涵的物理的和语义的内容进行计算机分析理解，以方便用户查询，其本质是对无序的多媒体数据流进行结构化，提取语义信息，保证多媒体内容的快速检索。多媒体信息的分析与检索流程如图 1.1 所示。

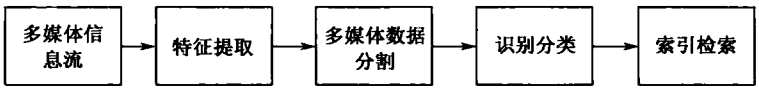

图 1.1　多媒体信息的分析与检索流程

特征提取是指寻找原始信号表达形式，提取出能代表原始信号的数据形式。与文本分析中的特征是关键字不同，多媒体数据中的特征可以是从图像与视频中提取的视觉特征，如色彩、纹理和运动等；也可以是从音频中提取的听觉特征，如音调、音质和音高等；还可以是从三维虚拟物体中提取的力矩和傅里叶因子等特征。所有这些提取出来的特征被用来表征多

媒体数据流，并将在后续的处理中被用到。

由于多媒体信息是时间序列数据流，例如，不能对 100 MB 大小的纯文本信息直接分析而要把它分成不同主题字段一样，也不能对几个小时长的视频或者音频直接处理，而是需要在其特征发生突变的地方进行分割，把连续多媒体数据流分成不同长度的数据片段，这是多媒体数据分割需要完成的任务；然后对分割好的数据片段进行处理。

连续的多媒体数据通过特征突变切分成不同的物理单元，需要分别对这些物理单元进行识别分类，将它们归属成事先定义好的不同语义类，这由多媒体识别分类这一步来完成。在这一步中，可以对分割出来的多媒体物理单元进行粗分，如将分出来的音频分类为静音、语音、环境音和音乐等，将切分出来的视频单元分类为屋外和屋内场景、体育新闻和广告节目等；也可以就某一事件或某一人物进行精细分类，如“爆炸”事件、“演讲”事件等。

多媒体检索的最后一步是对识别出来的语义建立索引，进行检索。

1.2.3 跨媒体信息定义与特性

传统的图像编码是按照像素点实现的，例如，所有的像素点汇集在一起就构成了整幅图像；音频编码也是一样的，所有采样点的汇总就构成了每个人最后听到的声音。而人们对图像或声音的理解并不是基于像素点，而是基于对对象（或场景）的理解。如果一张图片保存的是一辆汽车和一只狗，人们不会关心哪些像素点分别构成了狗和汽车的图案，而是直接形成了汽车和狗这两个对象的概念；对音频的理解也类似。在这种情况下，多媒体视频和音频编码方式与人脑的视频、音频场景产生机制发生了冲突，这个冲突也是目前在多媒体检索识别中尚未克服的一个困难，即多媒体底层视觉、听觉特征和多媒体高层语义之间存在一个“鸿沟”[6]。

跨媒体信息是指多媒体提取出的、能够跨越媒体类型的信息描述，一般来讲就是高层的语义信息描述，由于其与媒体类型无关，因此不同于传统多媒体检索方法中使用的底层物理特征，它更接近于人们的感知，符合计算机识别的最终目的，即使得计算机能够有人类一样的识别和检索能力。如何填补底层特征和高层语义特征（即跨媒体信息）的差异是多媒体检索中最具挑战性的课题。

1.3 信息融合与分析

融合分析是多媒体内容分析与理解领域的研究热点[9]，其出发点是：只有不同特征的融合才能表示多媒体数据所蕴涵的完整语义信息。融合分析与跨媒体检索思想的共同之处在于分析不同属性的特征以理解其表达的共同语义。目前的融合分析方法主要应用于单一类型多媒体数据的检索，大部分的单模态检索方法和系统都是通过提交一种查询例子，返回与其相

似的相同类型的多媒体对象。这些领域的研究通过提取多媒体数据相应的视觉或听觉特征，如颜色、纹理、运动、形状、短时能量和音调等，并将多媒体对象用底层特征构成的特征向量来表达，以实现对多媒体信息的管理和查询。

考虑到多媒体数据所表示的音、形、意等丰富信息，如果仅仅是单独使用视觉或听觉特征对音频、视频、图像进行分析，将导致部分多媒体信息丢失。基于内容的多媒体检索面临的“语义鸿沟”促使研究者开始通过对数据进行融合分析来提高语义理解的准确性[10]。

总体而言，多媒体融合与分析研究的分支和主要应用领域可以归纳为图 1.2 所示的范畴。

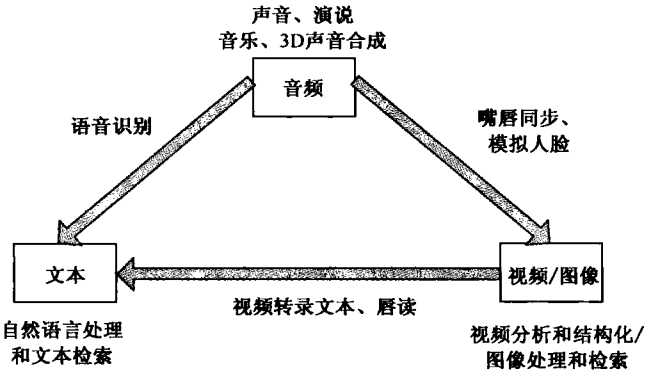

图 1.2　多媒体融合分析

融合分析方法可以帮助理解多媒体内容，以缩小语义鸿沟、提高检索效率，从而成功应用于文本检索、图像检索、视频检索和音频检索等领域，并实现在一定应用背景下不同类型的多媒体数据之间的定向转换。例如，使用语音识别技术，通过分析语音信号波形，把一个人讲话的语音信号转变为一串文字输出；针对一串文字，动画模拟人脸讲出这串文字的视觉、听觉动作等；通过嘴唇拼读技术，观察发音器官而推知讲话内容，帮助听力有障碍的人进行语音理解；而嘴唇同步则是研究如何在网络传输速度不快和视频、音频传输速度不一致的情况下，保证音频和视频信号同步。

但是，这些研究始终没有解决不同类型的多媒体数据之间的相关性计算问题，即跨媒体的相关性度量。在跨媒体检索环境下，用户提交一种类型的多媒体对象作为查询例子，系统不但可以返回相同类型的相似对象，而且可以返回不同类型的多媒体对象，例如用图像例子来检索音频片断。

上述基于融合分析的单模态检索面临两大困难：一方面是特征选取的差异性会在很大程度上影响融合结果，例如，两幅语义上相近的图像可能在颜色特征上完全不相似，但在形状特征上非常相似，如果融合过程中颜色特征的权重较高，显然不能取得良好的融合结果；另一方面，不同类型的多媒体特征之间存在的关联信息（如 Web 链接）通常被看成另一种附加特征进行处理和融合，而没有充分利用这种关联信息，并且作为附加特征的关联信息在学习

过程中往往不能更新。

由于以上两方面的原因，很难在多媒体语义理解过程中有效完成信息的互补和增强，系统检索性能的提高受到限制。于是，许多研究者致力于多媒体数据之间的关联挖掘，以期更准确地理解底层特征所要表达的高层语义。虽然关联挖掘研究不是为了实现检索过程中不同类型多媒体数据之间的灵活跨越，但这些研究都利用了多媒体数据之间潜在的相关性，与本书的跨媒体检索有相似之处。

下面三个方面归纳了基于关联挖掘的多媒体检索研究现状。

1）交叉索引关系

当一种类型的内容特征可以较好地反映多媒体语义时，可以用来对多媒体数据进行标注。例如，在视频流的内容分析和语义理解过程中，如果采用视觉特征不能得到满意的检索结果，则可以根据听觉特征识别出来的结果为相应的视频流建立索引。以“人群在爆炸声中奔散”为例，显然识别“爆炸”声音比识别物体的色彩和纹理在爆炸过程中的视觉变化更加容易，因此，只要把“爆炸”这个音频事件识别出来，就可以用它来索引相应的“爆炸”视频流，即用音频特征去为视频数据建立索引。

此外，文献[11]提出了一种多特征统一表示的索引方法；文献[12]针对互联网图像搜索，提出了一种基于语义和底层特征统一表示的索引结构，以支持基于关键字和特征的检索。卡耐基梅隆大学的 Informedia 研究项目，综合利用了图像理解、语音识别和自然语言理解等相关领域知识，实现对多媒体数据的检索与概括（Summarization），该项目采用了视频拼接（Video Collager）技术，先从多媒体数据流中提取人物、时间、地点、主题、事件描述和其他元数据去表示多媒体数据流，然后按照事先定义的语义模板，通过视频拼接把多媒体信息中文字、视频和音频所对应的元数据链接起来[13]。

2）链接关联模型

文本检索和 Web 挖掘等方面的研究表明，数据之间的链接关系是一种对检索非常有用的信息资源[14,15,20,21]。文献[14]通过挖掘不同类型的数据集之间潜在的联系来增强聚类效果；文献[15]提出在图像与其标注文本之间传递相似度，通过迭代计算，将文本内部的聚类结果传递到图像数据集中，利用数据集之间的关联找到潜在的图像相似关系。类似地，文献[16]提出了通过挖掘 Web 图像和周围文本标注之间的关联信息来进一步学习 Web 图像的相似度匹配，其基本思想是：已知图像和文本之间对应的链接关系矩阵 $\boldsymbol{M}$，以 $\boldsymbol{M}$ 为桥梁，将文本集的相似度关系图向图像集传递，同时将图像集的相似度关系 O 向文本集传递，并控制传递过程中文本集对图像集的影响大于图像集对文本集的影响程度，使得修正后图像数据集的相似度匹配结果更加符合真实的语义关系，从而提高图像检索的查准率和查全率。

上述方法本质上是以文本和图像之间的关联信息为桥梁，用一种特征空间的数据集的拓扑结构去修正另一种类型的特征空间中数据集的分布，最终达到一个稳定的状态。这种方法可提高图像语义理解和检索的效率，其局限性是无法测量两种不同类型的多媒体数据之间的

相关性，即两种不同的特征空间之间的相关性度量问题无法解决，所以也就无法实现用一种类型的多媒体数据检索其他类型的相似多媒体对象。

3）多媒体关系图

使用图模型表达数据以及数据间的相互关系可以很好地将数据集结构化，并有效地发现潜在的数据关系。不同领域的研究工作都已经证明图模型在数据表达方面的有效性[17][18]。文献[19]提出将 Web 图像的底层内容、环绕文字以及图像与图像之间的链接看成三种不同类型的特征，然后根据欧氏距离分别为之构造 K-近邻图模型 W^{c}、W^{t} 和 W^{l}，将异构的 Web 数据用多媒体关系图的方式进行结构化处理。在此基础上，在用户提交关键字作为查询请求的时候，可以非监督式地进行聚类：首先选择环绕文字包括查询关键字的图像集 $I^{(0)}$，然后根据多媒体关系图找到与 $I^{(0)}$ 相邻接的图像数据集 $I_{adj}^{(0)}$，将 $I^{(0)}$ 和 $I_{adj}^{(0)}$ 作为初始查询结果返回给用户。由于 $I^{(0)}$ 和 $I_{adj}^{(0)}$ 本身都是未标注的图像集，因此这种方法可有效缓解监督式学习中的小样本问题。此外，在 Web 环境下如果仅仅根据文本进行检索很可能会丢失大量的正例，而 X. J. Wang 等人的方法直接采用了非标注样本，提高了查全率（Recall）。

1.4 跨媒体信息检索

互联网的发展带来了更多的多媒体应用，随着多媒体网站的日益发展，针对跨媒体信息的检索和提取技术越来越引起人们的关注。然而到目前为止，这方面的研究工作还处于发展时期，大多数的信息提取和检索方法都是采用文字检索或人工标注来进行的，将较为复杂的跨媒体信息转移到较为成熟的文本信息上。这种方法在多媒体个数较少时可以较好地得到应用，然而当多媒体的个数不断膨胀时，人工标注就将很难胜任；而文字检索多采用上下文检索，也很难取得令人满意的效果。

跨媒体信息反映的是高层的语义特征和信息，前面也提到，它和底层特征有着较大的差距，因此，很难有统一的标准来表示跨媒体信息。但是我们注意到，实际应用中并不在意跨媒体信息的具体的值，而在意的是两个多媒体的跨媒体信息的差值。因为在检索过程中，就是要找到离目标多媒体最近的若干多媒体作为返回值。因此，可以使用相似度度量矩阵来表示数据库中任意两个多媒体之间的距离。对于数据库外的多媒体，一般可利用底层特征找到同类型多媒体的若干近邻，再以这些近邻作为出发点查找最近的多媒体。

事实上，单纯使用计算机来进行跨媒体信息的检索存在很大的困难，目前研究比较深入的一种方法是反馈机制。反馈机制考虑到了人的因素，通过用户的判断来使系统更加精确。可以说，在找到一种完全的、靠计算机进行的检索方法之前，用户的反馈结果能使系统性能有很大的提升。

试读结束：需要全本请在线购买： www.ertongbook.com