

医学影像数据库设计及图像检索技术研究

胡 军

1999 年 5 月

西北大学计算机科学系 710069

摘 要

医学影像数据库是建立在图像数据库、图像处理与理解、人工智能、远程与局部网络技术以及医学领域知识基础上,支持医学影像数据有效存储、传输、检索和管理的数据库技术。我们的目标是充分考虑国内医学影像管理和处理的需要以及医疗设备现状,研究并开发出既适应电子病历管理,又能为医学多媒体教学、虚拟手术、手术计划与导航、远程诊疗等应用提供数据管理与服务的核心数据库系统。

本文首先在比较和分析各类图像检索系统的基础上,提出了一个用于图像数据库系统的图像检索模型,这个模型不仅可以用来作为描述和比较图像检索方法的一般框架,也可以作为实现图像检索机制的参考结构。然后在医学影像管理需求和图像检索模型的基础上设计了三层 Client/Server 结构的医学影像数据库应用系统的系统结构,并进行了基于关系模式的数据库设计。针对远程图像存取时网络的有限带宽问题,本文从软件的设计与实现方面提出了一些相应的对策。同时医学影像数据库系统中有机地集成一些图像处理功能也是必要的,本文对医学影像涉及到的格式转换、数据压缩和处理方面的部分问题进行了分析和研究,并给出了一个用于图像数据库系统的基于构件的嵌入式图像处理系统模型和结构。

医学影像数据库研究中一个基础的、重要的、有意义的工作是图像数据库的基于内容的检索技术研究。本文提出了一个通过直方图和二值集方法解决 CBIR 问题的有效的技术路线,使得我们能以一种一致的、紧凑的、自然的、完整的方法来看待和表达图像的颜色、纹理和形状特征,通过四叉树特征分解有效地表达图像中对象的空间关系。对此本文就基于颜色的图像检索进行了详细的介绍。

为提高数据库的可用性并拓宽信息的服务范围,针对医学影像数据的远程访问需要,我们对医学图像数据库的 Web 检索服务进行了研究和实践。

医学影像信息管理的研究,不仅有着理论的意义,也将对实际的医院管理工作起到重要的作用,对现有的医疗现代化带来革命化的影响,有着良好的社会效益和很大的经济效益。

关键词 图像数据库 医学影像 图像检索模型 图像处理 基于内容的图像检索
三层 Client/Server 结构 Web/Server

CONTENT-BASED RETRIEVAL ON A MEDICAL IMAGE DATABASE

Hu Jun

May, 1999

Department of Computer Science, Northwest University

Abstract

Medical image database, which is based on image database, image processing & comprehension, artificial intelligence, remote & local network, including knowledge of medical, should support medical image data access, transmission, retrieval and manipulation effectively. In China, there is a number of special features on medical apparatus and requirements of medical image management and processing. Taking them into account, our study motivation focus on a database system kernel, which is suitable for management of electronic case, multimedia medical education, virtual operation, operation planning and guidance as well as remote diagnosis and treat.

This thesis presents an image retrieval model for image database firstly. It can be referred to as framework for retrieval methods description and comparison, or architecture for image retrieval systems. Based on that, considering the requirements of medical image management, then we present the architecture design of 3-Tier Client/Server for medical image database application, as well as the database design of relation schema. Due to the limited bandwidth of the network that would jam image access, some countermeasures are suggested in software design and implementation. Furthermore, it is necessary for image database to integrate some important and useful functions of image processing, we herewith make some efforts to discuss the issues in format conversion, data compression and processing which is medical image database system concerned. It would be concluded with an embed image processing system for image database, which is component based.

A fundamental and significant work in medical image database is the study of content based image retrieval (CBIR) technique. This thesis present a technical line based on histogram and binary set for CBIR, which is effective, uniform, compact, natural and complete for color, texture and shape feature representation, and on the other hand, the spatial relation among the objects is represented by quadtree based decomposition. As a demonstration, color based image retrieval is discussed in detail.

In order to improve the usability of image database and broaden the information service, considering the requirement of remote access to the medical image database, web based retrieval service is also studied and put to the test.

The study in medical image management is significant not only in theory, but also in practice. The social and economic effect on hospital management and medical treatment modernization would be revolutionary and far-reaching.

Keywords image database, medical image, image retrieval model, image processing, content based image retrieval, 3-tier client/server, web/server

目 录

第一章 绪论	1
§1.1 研究背景.....	1
§1.2 图像数据库.....	1
§1.3 医学影像数据库.....	2
§1.4 论文的主要工作和意义.....	3
第二章 图像检索系统分析	5
§2.1 图像检索方法的分类.....	5
2.1.1 直接查询(Direct Query)	5
2.1.2 图例查询(Query by Pictorial Example)	7
2.1.3 浏览(Browsing)	9
2.1.4 综合检索方法.....	11
§2.2 图像检索模型.....	12
2.2.1 Infoscope.....	12
2.2.2 Grosky	13
图像检索模型.....	14
§2.3 用图像检索模型分析检索方法.....	16
2.3.1 直接查询 (Direct Query)	16
2.3.2 图例查询 (Query By Pictorial Example)	16
2.3.3 浏览 (Browsing)	17
§2.4 小结.....	18
第三章 医学影像数据库系统的设计	19
§3.1 三层 Client/Server 结构	19
3.1.1 三层 Client/Server 的基本硬件结构.....	19
3.1.2 三层 Client/Server 的功能	20
3.1.3 三层 Client/Server 结构的优点	21
3.1.4 中间件.....	21
§3.2 系统结构设计：三层结构的图像检索模型	22
§3.3 数据库设计.....	23
3.3.1 图像库.....	23
3.3.2 特征库.....	24
3.3.3 联接库.....	24
3.3.4 领域库.....	25
§3.4 远程图像存取中的问题与对策.....	27
3.4.1 问题.....	27
3.4.2 多分辨率的检索结果.....	27
3.4.3 压缩存储.....	27
3.4.4 切分传送.....	27
3.4.5 检索结果规模控制.....	28
3.4.6 在表示层实现部分图像分析	28

3.4.7 缓冲技术.....	29
§3.5 开发与运行环境.....	29
3.5.1 开发环境.....	29
3.5.2 目标系统运行环境.....	29
§3.6 小结.....	30
第四章 医学影像数据库中的图像处理.....	31
§4.1 医学影像的格式.....	31
§4.2 医学影像的压缩.....	32
4.2.1 无损压缩技术.....	33
4.2.2 图像有损编码.....	33
4.2.3 在医学影像上的压缩实现.....	33
§4.3 医学影像的处理.....	33
4.3.1 颜色调整.....	34
4.3.2 图像锐化.....	34
4.3.3 图像增强.....	34
4.3.4 图像均衡与直方图调整.....	35
4.3.5 缩放.....	36
4.3.6 其他处理.....	36
§4.4 基于构件的嵌入式处理系统设计.....	37
§4.5 小结.....	38
第五章 基于内容的图像检索技术 (CBIR)	40
§5.1 CBIR 的技术特点	40
§5.2 CBIR 的一般过程	40
§5.3 一个解决 CBIR 问题的总体技术路线.....	41
5.3.1 基于二值直方图的特征表达.....	41
5.3.2 基于四叉树的空间关系表达.....	42
§5.4 基于颜色特征的检索.....	43
§5.5 基于纹理特征的检索.....	43
§5.6 基于形状特征的检索.....	45
§5.7 检索效果评价方法.....	46
§5.8 小结.....	47
第六章 基于颜色特征的图像检索.....	48
§6.1 特征表达.....	48
§6.2 一些相关工作和比较.....	48
§6.3 彩色视觉的色度学和彩色感觉.....	49
§6.4 颜色空间的变换和量化.....	49
6.4.1 颜色变换.....	50
6.4.2 颜色量化.....	50
§6.5 颜色空间.....	51
6.5.1 RGB 颜色空间	51
6.5.2 线性变换颜色空间.....	51
6.5.3 Munsell 颜色表系统.....	53

6.5.4 CIE 颜色空间	53
6.5.5 HSV 颜色空间.....	54
§6.6 表达颜色特征.....	56
6.6.1 颜色直方图.....	56
6.6.2 颜色直方图的性质.....	57
6.6.3 二值颜色集.....	57
6.6.4 颜色直方图压缩.....	58
6.6.5 二值集快速映射 (BSFM)	59
§6.7 直方图的归一化与度量关系.....	59
§6.8 度量方法.....	60
6.8.1 弱相似性距.....	60
6.8.2 相似性距.....	63
§6.9 实验效果.....	65
§6.10 小结.....	66
第七章 基于 WEB 的图像检索	67
§7.1 客户端直接访问数据库.....	67
§7.2 基于中间件的 Web 数据库服务.....	68
7.2.1 CGI 技术.....	68
7.2.2 API 技术	69
7.2.3 Web_DB 技术产品	69
7.2.4 Cold Fusion.....	69
§7.3 Web/Server 模式的主要问题与对策	70
§7.4 基于 Web 的图像检索	71
7.4.1 几种流行的图像搜索引擎.....	71
7.4.2 Image Surfer 的实现技术.....	72
7.4.3 值得医学影像 Web 检索借鉴的技术.....	72
§7.5 小结.....	73
第八章 总结及展望.....	74
参考文献.....	77
致 谢.....	79

第一章 绪论

§ 1.1 研究背景

随着近年来人们利用光、声、电、磁等各种物理能的作用机制,配合化学染色剂或造影剂,借助计算机软硬件技术,促使了医学诊断可视化技术的长足的进步,医学影像信息出现爆炸性的发展。医学影像技术在医学科研与教学、临床诊断、远程诊疗中均有极其重要的作用,现代医学离不开医学影像信息的支持。如何高效、方便的对大量医学影像进行智能化处理和管理,是提高医学研究水平、实现医院管理现代化的关键所在,也为计算机应用技术的发展提出了新的研究课题。

近年来,国内和国际上有部分学者就图像信息的存储、传输和处理作了大量的研究,尤其是最近几年关于图像数据库的研究取得了可喜的进展,如 IBM 公司的 QBIC 系统、美国哥伦比亚大学的 VisualSEEK 系统等,他们的工作为开展医学影像数据库的研究打下了坚实的基础。但这些研究工作落实到医学影像上还是有很大差别的,对医学影像数据库的特殊性没有予以充分的考虑。医学影像数据的管理和操纵的技术,有别于其他影像数据。例如,其数据的器官结构形状、生物纹理特征、医学领域的相关知识和临床医学对影像信息的需求等。建立医学影像数据库,可借助现有图象数据库和工程数据库的研究基础,也必须充分利用医学数据的相关知识。在此基础上,综合利用生物医学工程等多学科的新技术新方法,提供医学影像数据的集成管理,它将是病情诊断、手术计划与导航、远程诊疗、多媒体教学等系统的核心和基础,对诊断、治疗、教学起到十分重要的作用。这种系统是医院急需的高科技软件,目前还未见有成熟的商业软件产品。

自从 1993 年以来,西北大学可视化技术研究所与第四军医大学、陕西省人民医院、西安医学院附属医院的医务人员合作,在医学图象的智能化处理和可视化技术研究方面做了大量的工作,完成了“三维医学可视化分析系统”(陕西省重点科技攻关项目,已通过陕西省科委组织的鉴定,文献[25][26]),“电子病历管理”(已在第四军医大学西京医院应用),在法医人类学的面貌复原研究工作上(国家“九五”重点科技攻关项目,文献[28])取得了可喜的进展(已通过国家科委和公安部组织的中期检查),计划采集医学图象数据 20000 多幅,目前已入库 7000 余幅。以上两项研究工作已解决了医学图象数据的采集、医学图象数据的压缩存储、医学图象的可视化技术等关键问题,医学图象的智能化管理已成为迫切需要解决、而且也已具有解决基础条件的问题。

§ 1.2 图像数据库

广义说来,所谓的图像数据库就是能够有效存储、管理、组织和检索图像数据的数据库。但图像数据库不仅仅只是现有数据库技术在图像数据上的一种应用,因为图像作为一种多媒体信息,还含有对人类感觉系统(视觉与听觉)比较明确而在数据中未有效描述的内容线索,基于内容的检索符合人类感觉经验,易于接受,也成为图像数据库的一个重要特征。

对于多媒体来说,每一种媒体数据都有一些难以用字符和数字符号描述的内容线索,如图像中某一对象的形状、颜色和纹理,视频中的运动,声音的音调等。当用户要利用这些线索对数据进行检索时,就不得不首先将其人工转化为文本或关键词形式。但这种转换带有一

定的主观性,且极其费时。现在,网络和数据库中的多媒体数据量非常庞大,人们在应用中不仅要求数据库及其它信息系统能对图像、视频和声音等媒体进行存储以及基于关键字的检索,而且要求对多媒体数据内容进行自动语义分析、表达和检索。所谓基于内容的检索(Content Based Retrieval,CBR)就是指根据媒体和媒体对象的内容语义及上下文联系进行检索。

基于内容的检索需要利用图像处理、模式识别、计算机视觉、图像理解等学科中的一些方法作为部分基础技术。CBR 不仅是基于内容,而且是一种信息检索技术。它从认知科学、用户模型、图像处理、模式识别、知识库系统、计算机图形学、数据库管理系统,以及信息检索等领域中获得启发,引入新的媒体数据表示和数据模型,产生出有效、可靠的查询处理算法和可视化查询接口,以及与领域无关的检索技术和系统结构。

CBR 与模式识别、图像理解及计算机视觉等学科的重要区别在于, CBR 是一种信息检索技术,应能从大型分布数据库(集)中,以用户可以接受的响应时间,查询到所要求的信息。

基于内容的检索是信息检索中的一种技术,它主要利用媒体的视觉和听觉特征进行检索。在很多应用中,基于关键字的检索技术是行之有效的检索方法,因此,完整有效的信息查询和检索系统应该包括常规的基于客观属性(关键字等)的检索、基于内容的检索、对象关联检索及在这些检索之上的概念查询。

对象关联检索用于多媒体复合对象和超媒体文献检索,它以媒体对象之间的时空关系和超媒体对象之间的关联为线索进行检索,检索结果可能是满足查询要求的关联所构成的语义网络。概念查询以这些检索技术为基础,针对用户提出的查询概念要求,对多媒体数据库(集)进行检索。

CBR 技术是一项实用的技术,主要应用于以下几个方面:将基于内容检索引擎嵌入到常规数据库管理系统中,以实现多媒体数据的检索;在信息检索系统中,对专用领域的视频、图像和文档库进行检索;对 Internet 上包含在 Web 信息网中 HTML 页面上的多媒体数据进行基于内容检索等。

基于内容的图像检索(CBIR)主要依据图像的颜色、纹理、形状特征,以及图像中子图像(目标,又称对象)的特征进行检索。主要包括:

颜色查询 使用户查到与用户所选择的颜色相似的图像。

纹理查询 使用户查到含有相似纹理的图像。

形状查询 用户选择某一形状或勾勒一幅草图,利用形状特征(主轴方向、矩、内角特征等)或匹配主要边界进行检索。

图像对象查询 对图像中所包含的静态子对象进行查询。查询条件可以综合利用颜色、纹理、形状特征、空间关系、逻辑特征和客观属性等。

§ 1.3 医学影像数据库

医学影像数据库是建立在图像数据库、图像处理与理解、人工智能、远程与局部网络技术以及医学领域知识基础上,支持医学影像数据有效存储、传输、检索和管理的数据库技术。我们的目标是充分考虑国内医学影像管理和处理的需要以及医疗设备现状,研究并开发出既适应电子病历管理,又能为医学多媒体教学、虚拟手术、手术计划与导航、远程诊疗等应用提供数据管理与服务的核心数据库系统。它既不同于传统的数据库,也不同于一般的图像数据库,在需求和实现上需要研究的内容和要解决的问题包括:

(1) 医学影像数据库的检索

医学影像数据库需要传统的基于关键词的检索方式,而且也将是主要的检索方式。这就需要充分了解医学影像的描述信息,按照医学的知识结构建立有效的数据模型。但一般来说,这种检索技术往往不能满足人们的需要。由于医学影像内容具有丰富的内涵,在许多情况下仅用几个关键词难以充分描述人们对影像检索的需要,而且作为关键词的图像医学特征的选取也有很大的主观性。这样基于内容的图像检索(CBIR)技术可以得到充分应用,融合图像理解技术,从而可以提供一种从巨容的图像库中,根据人们提出的要求进行有效检索的方法。基于内容的检索技术应用在医学影像数据库上,不仅需要一般的图像学特征,而且还需要将这些特征与医学需要相结合,将人工智能技术与图像技术相结合,建立附加的医学知识库,在知识库的导引下进行符合医学常规的内容检索。从而提出了新的研究方向,如:

在自动的或计算机辅助的器官分割技术基础上进行的基于形状的检索;

基于器官或病变纹理的检索;

在基于颜色的检索上充分考虑各种医疗设备对相同器官和组织表现差异;

(2) 医学图象的智能化处理

医学图象的智能化处理在医学影像数据库的研究与实现中有着重要的作用和地位。医学影像数据库系统出于特征抽象的需要,必须在系统内部有效地提供无缝的图像分析与处理支持;医学影像数据库系统的用户都不是图像处理专家,往往不能有效地使用复杂的通用图像处理系统,他们只需要一些常用的适合医学影像的处理手段和技术;医学影像往往需要进行一些与应用和领域有关的特殊分析与处理,而且用户没有必要为了完成简单的图像处理工作在数据库系统和图像处理系统之间来回的切换,对医学影像进行格式转换、压缩和处理分析;尤其在网络环境下,将图像处理功能集成到不同层次的数据库系统服务上,可以根据需要进行“就地”处理,而没有必要都传送到客户应用端处理,可以有效地降低传输开销。

所以对医学影像数据库中的图像处理的任务和特点进行研究,并给出具有较高可用性、适应性和效率的图像处理模块的设计,也是医学影像数据库研究的重要内容之一。

(3) 数据库与医学影像设备的接口

1) 应提供针对医学影像特征的的用户需求接口描述和协议;

2) 提供针对多种医学影像格式的统一数据转换接口与医学图象的数据采集接口。进行适应性的数据输入。在对已有的CT机,核磁共振图象(MRI)机器的接口基础上,进一步扩大和推广接口的适应能力。

(4) 医学影像数据库的三维可视化

在已完成的“三维医学可视化分析系统”的基础上,实现医学数据库的可视化接口。提供对该数据库的信息进行二维上编辑处理的功能,并可对于库中的医学体数据,亦可实施三维观察与操纵(文献[27])。

(5) 医学影像数据库的远程访问

医学影像往往具有分布式存放的特性,应充分考虑远程教学与诊断、甚至计算机协同工作的需要,提供局部和远程尤其是Internet访问支持,可以考虑Client/Server、Web/Server等数据库访问方式。尤其需要注意的是如何建立高效的医学影像的数据远程访问服务,将是本题目不同于一般Internet数据库服务的重要研究课题。

§ 1.4 论文的主要工作和意义

针对医学影像数据库的需求和其中涉及的一些问题,我们从1997年下半年开始就开始关注在图像数据库方面国际和国内的最新成果,并着手对有关技术和问题进行研究,现已取得了一些进展。与本文有关的主要工作有:

(1) 在比较和分析各类图像检索系统的基础上, 我们提出了一个用于图像数据库系统的图像检索模型 (Image Retrieval Model), 这个模型不仅可以用来作为描述和比较图像检索方法的一般框架, 也可以作为实现图像检索机制的参考结构 (第二章)。

(2) 在医学影像管理需求和图像检索模型基础上我们设计了三层 Client/Server 结构的医学影像数据库应用系统的系统结构, 并进行了基于关系模式的数据库的设计。针对远程图像存取时网络的有限带宽问题, 我们从软件的设计与实现方面提出了一些相应的对策 (第三章)。

(3) 医学影像数据库系统中有机地集成一些图像处理功能是必要的, 我们对医学影像涉及到的格式转换、数据压缩和处理方面的部分问题进行了分析和研究, 并给出了一个用于图像数据库系统的基于构件的嵌入式图像处理系统模型和结构 (第四章)。

(4) 我们提出了一个通过直方图和二值集方法解决 CBIR 问题的有效的技术路线, 使得我们能以一种一致的、紧凑的、自然的、完整的方法来看待和表达图像的颜色、纹理和形状特征, 通过四叉树特征分解有效地表达图像中对象的空间关系 (第五章)。对此本文就基于颜色的检索进行了详细的介绍 (第六章)。

(5) 为提高数据库的可用性并拓宽信息的服务范围, 针对医学影像数据的远程访问需要, 我们对医学图像数据库的 Web 检索服务进行了研究和实践。

(6) 根据上述研究和工作的成果, 我们实现了医学影像数据库系统的原型系统。

医学影像的智能化处理与管理技术在医学科研与教学, 以及临床诊断方面有着广阔的前景, 有许多需要我们计算机科研人员去深入研究的课题。现代医学科学研究和临床诊断所需要的医学影像是多种多样的, 例如: 病理切片图像、X 射线透视图像、X-CT 断层扫描图像、 γ 射线的核医学影像、电磁波的磁共振 (MRI) 扫描图像、通用 B 型超声波扫描图像、彩色多谱勒超声波扫描图像、红外线的热成像仪图像, 以及各种电子内窥镜图像等等。这些医学影象信息的智能化管理研究, 不仅有着理论的意义, 也对实际的医院管理工作起到极大的作用。医学影象信息的计算机管理, 包括影象数据的采集、电子影象病历、影象数据的智能处理、远程诊疗系统等, 这些每一个都将给现有的医疗现代化带来革命化的影响, 有着良好的社会效益和很大的经济效益。

目前我们业已完成的研究和工作, 不仅使得我们能够有机会在的图像数据库的研究领域占领一席之地, 也使得医学影像数据库的广泛应用和推广有了扎实的工作基础。

第二章 图像检索系统分析

从本质上来说,医学影像数据库是一种医学领域的图像数据库。图像数据库不同于一般数据的一个重要特点和任务便是图像的检索。

由于在获取和存储图像方面的成本的降低,目前存在着大量巨大的、而且是分布的图像集合。在一个给定的大图像集(比如超过 10,000 幅图像)中检索特定的图像是很困难的。一幅一幅地浏览显然是不现实的。为此,目前有一些方法辅助人们检索图像,但都各有其优缺点及适用性。本文首先介绍了以大型图像数据库数据检索为目标,先后出现的七种共三大类图像检索方法(Image Retrieval Methods)。随后定义并建立了图像检索模型(Image Retrieval Model),并在该模型的基础上分析比较各种图像检索方法的特点。

§ 2.1 图像检索方法的分类

大体上来说,图像检索系统的发展经历了两个阶段:关联信息检索(Based on data about the images)和图像数据检索(Based on the image data)。在图像检索技术的发展过程中,文本检索技术的迅速发展一方面促进了图像检索技术的发展,有很多有效的技术和方法应用到图像检索中;另一方面也使得图像检索技术的发展前期基本上照搬了文本检索的一些方法,从而影响了图像检索技术的发展。这里首先介绍并讨论七种图像检索方法的主要思想和主要差别。这些方法的特性将在本章第三节中用图像检索模型加以分析和讨论。

2.1.1 直接查询(Direct Query)

传统的数据库系统大都提供了这种方法用于检索图像。直接查询是一种用结构化的相关的非图像信息来检索图像的方法。

传统的数据库具有下面的两个特点:

- 在数据结构上,要求检索所提供线索的数据类型与数据库中的检索对象的相应域的数据类型一致。在传统数据库的图像检索中,用户基本上没有手段为此提供原始的图像数据线索。
- 直接检索通过检索条件将集合精确地划分为匹配与不匹配两个子集。用户能为数值型的数据提供形如“ $0 < x < 10$ ”的检索条件,也能为字符型的数据提供形如“ $'a' < x < 'd'$ ”的条件,虽然传统数据库提供了存储图像数据的手段,如 BLOB(Binary Large Object)或 General 等数据类型,却并未提供描述图像数据条件范围的手段,也即图像数据对象的表达不象数值、字符等一般数据对象在属性和操作上比较完备,所以用户不能为检索图像数据描述一个意向范围。

正是由于这两个特点,传统数据库不能直接处理图像数据的检索需求,而只能通过与图像数据相关的结构化信息进行。结构化信息是对客观对象有目的有组织的抽象,例如人员数据库中的姓名和年龄等信息,可以通过这些信息来检索有关人员的图像信息,因为数据库中的元组抽象了有关人员的必要的足够的信息。值得注意的是,在这里图像数据也是从客观对象抽象而来,是一种结构化的、数据库能够以一定格式存取的数字化的图像数据(Data),之所以不认为它是一种结构化信息(Information),是因为其中没有任何语义(Semantics)引入。

下面将讨论两种形式的直接检索方法。一种是根据对图像的描述来检索图像,另一种

是通过图像分析技术分析出的图像特征来检索图像。

2.1.1.1 根据描述(Descriptions)查询

抽象图像最显而易见的手段是用文字描述，这种抽象方法常称之为描述或注解。基于描述的查询是一种仅根据用户定义或描述的特征值查询图像的检索方法。用户定义的特征值常与插入图像时的应用环境有关，如存储者的知识、图像的标题等。

比如，一幅小汽车的图像可以根据小汽车的颜色或生产厂家进行索引。用户在查询时可以直接使用数据库查询语言如 SQL 表达其查询需求。查询中的条件可以使用图像的描述，而查询的结果将由原始图像数据组成，如所有有关黑色桑塔纳轿车的图片。

在这种情况下，数据库系统除了需要提供存取图像的工具和手段外，并不需要额外的功能支持这种直接查询。很多商用的数据库系统都支持这种查询方法，如 Microsoft Access、Borland Paradox、Sybase Anywhere 等，这些系统通常用 BLOB (Binary Large Objects) 类型存取图像。

根据描述进行的直接查询需要大量的描述。对大规模的图像集合而言手工描述并索引在实际工作中是不可行的，除非存在对其他目的而言已经可用的描述，比如在邮购目录中产品已经被详细描述并附有图片，由于此时描述和图片都是同一个对象的抽象，所以描述可以用来对图像进行有效的检索。

另一个有用的高级技术是根据图像所在的上下文来产生描述。文献[16]介绍了一个名为 Piction 的系统，该系统借助于视觉语义和自然语言处理，从报纸的照片中识别人脸。Piction 通过图片的标题将一个短语与照片甚至照片的一部分相联系，描述和索引过程是由计算机系统自动完成的。图像的检索则完全根据描述进行，图像与图像之间不再作任何比较。

给所有图像赋予适当关键字的目标是可能达到的，而描述的含糊性则可能是很难克服的问题。自然语言处理中的一些技术如辞典技术，可以作为降低含糊性的辅助手段。即便如此，如果一幅图片在不同的上下文环境中可能被赋予完全不同的关键字。如图 2-1，可能被认为是一个器官、苍蝇的头部、四季豆或一架电话等。

文献[17]指出，“观察图像者常常能为图像给出更合适的描述关键字，而不是图像的作者”，但是，我们认为，描述只表达了存储者对图像的认识和观点，并不能反映检索者的对图像的看法，而且与检索者的检索目的有关。

由此可见，基于描述的直接查询的两个主要的缺陷是，需要对所有图像进行艰苦的描述，而描述又具有含糊性。这个方法的主要优点是检索是基于图像的语义进行的，在每个对象具有确定的或唯一的描述时，根据描述查询是非常有效的。

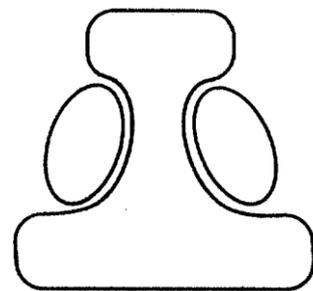


图 2-1 一幅器官的图片？

2.1.1.2 根据图像特征(Image Features)查询

这里的特征指的是图像的视觉特性。基于图像特征的查询是根据直接从原始图像数据计算得到的特征值进行查询的一种检索方法。图像特征是从图像抽象出的、计算机能够处理的数值或二进制串，是图像的非信息属性 (Non-information-bearing Attribute)。

特征可以是简单的度量信息，如图像中红色像素的数量、比重，也可以是图像中对象的性质，如对象的形状、位置等。传统的数据库无需作任何扩展就可以用来存取用数值表达的特征，但对复杂的应用，数据库需要扩展的数据类型和数据操作以支持几何及拓扑对象的表达。

当然这种方法也有一些问题。比如要从小汽车的图像库中找出所有关于红色小汽车的图片，假定数据库用字段“red_pixels”的值表示图像中红色像素的百分比，我们可以构造

相应 SQL 查询:

```
SELECT pictures FROM database
WHERE red_pixels > :lowerbound ;
```

查询是以阈值为基础的,并不复杂,关键是用户必须给出合适的参数,即便如此查询结果也常常出人意料。用户必须精心选择阈值参数,使得阈值既要足够大以避免图片上过小的红色区域,也要足够小以检索出含有很小的红色小汽车的图片。如果阈值太小,在车窗上贴有红色标签的褐色汽车可能被检索出来;如果阈值太大,如果图片中的红色小汽车所占的面积小于阈值将被忽略,用户将处于两难的境地。而且,上述的讨论也是有前提的,也就是数据库中所有的图片都是关于小汽车的,且每幅图片中小汽车都是主要的场景。如果库中含有其他图片,或小汽车的背景过于复杂或强烈,上述的查询的命中率将更加低下。

基于图像特征的查询的主要优点是很有可能对图像特征进行自动抽取。一旦建立了抽取例程,将不再需要人为的干预。与人工处理相比,计算机处理无论在时间和经济上都具有很大的优势。

而且,预定义模块的使用使得图像的抽象更加客观。如前所述,为了对图像中的每幅图像作出注解需要一组工作人员的艰苦工作,而每个成员在解释图像时都或多或少地带有一些主观臆断。用一个自动特征抽取器来取代这些工作人员进行特征抽取与描述,其看待图片的观点是恒定而且客观的。但在另一方面,人工描述要比自动抽取更具灵活性。自动特征抽取器的设计者必须考虑到图像所有可能的特征内容,而人工描述可以使用与生俱来的复杂的评价模型并结合毕生的视觉经验对图像作出描述,这又是自动特征抽取难以达到的。

综上所述,根据图像特征的查询的主要优点是特征抽象过程可以自动地、客观地进行。其缺陷在于,由于查询中用户必须给出含有图像特征值的条件表达式以构造查询,对用户来说是相当困难且难于理解的。为了降低图像特征抽取模块的复杂度,使用图像特征的检索系统的应用常常限制在特定的领域。

2.1.2 图例查询(Query by Pictorial Example)

图像特征的使用在直接查询的情况下看上去似乎不那么富有成效,但不能说明图像特征的引入是错误的。事实上,当前的大多是图像检索技术都使用了图像特征。在用于文字与数字信息检索的 QBE (Query by Example) 概念的基础上,文献[14]早在 1980 年就提出了图例查询的概念。图例查询是一种由一幅或多幅图像实例构造查询的图像检索技术。

QBE 是 1975 年由 M.M. Zloof 提出,1978 年在 IBM370 上实现的关系数据库语言。QBE 是一个很有特色的语言。首先,用户通过显示终端使用 QBE,不仅查询的结果用表格的形式显示,而且用户的查询要求也是在终端上用填写表格的方式给出的。它具有直观和对话的特点。其次,因为 QBE 是在显示屏幕的表格上进行查询,所以它具有“二维语法”的特点,而其他语言的语法则线性的。第三,QBE 的字面含义是“用例子进行查询”,这种操作方式用户容易掌握,特别是缺乏计算机和数学基础的非程序员用户乐于接受。Thomas 等人曾作过实验表明 QBE 的用户性能最佳,最易学习、掌握(文献[7])。与 SQL 之类的查询语言相比,用户在构造查询时 QBE 更具透明性。同时也需要指出的是,QBE 不适合构造复杂的和嵌套的查询。

前面我们曾指出了在图像检索中精确匹配的查询方法存在的问题。作为图例查询方法的一部分,文献[14]同时也指出图例查询的方法需要相似性度量算子。除了需要从原始图像数据中抽象特征外,用来衡量特征抽象数据的匹配程度的方法也是必须的。针对精确匹配的问题,相似匹配的引入使得数据库的划分转变为相似性的对比。

文献[17]描述了对非结构化的文档进行的全文信息相似性检索与一般数据库的数据检索的差别。表 2-1 反映了这两种检索方式的主要区别。这个表同样可以不加修改地应用在图

像检索上。数据检索相当于直接检索，而信息检索则可以看作图例查询和近似匹配。

	数据检索	信息检索
存储信息的表达	定义良好的对象和数据	非结构化的信息
响应需求的方式	直接	提供用户可能感兴趣的结果
用户对结果的满意度	满意或不满意	一定的满意程度
系统查询成功的标志	用户是否找到了需要的图像？	系统是否最大可能的满足了用户的需要？

表 2-1 数据检索 vs 信息检索

从表 2-1 可以看到，信息检索是基于可能性的。当用户查询一个文档集合时，信息检索系统对集合中的每个文本返回一个所谓相关度的数字。这个数字反映了文本与查询匹配的程度。比如，如果用户需要查询包含“西北大学”的文档，那么“西北大学”在其中出现了 10 次的文档比只出现了 3 次的文档显然具有更高的相关度。对相关度进行分级，系统根据相关度的高低次序提交所有文档的列表供用户选择。用户可以从列表的顶端选择到与其查询最有可能相适应的文档。

图例查询尽管在用来计算相关度的方法上与全文信息检索有很多的区别，但两者使用了相同的原则来满足用户的信息需求。用户可以向图例查询系统提供一幅红色小汽车的示例图片，要求系统查找看上去相像的所有图片。系统将计算示例图片的特征值并与数据库中图片对应的值进行比较。这里我们再一次使用“red_pixels”属性作例子，一个贴着红色标签的蓝色小汽车将比不含任何红色像素的蓝色小汽车获得更高的相似级别。无论如何，在图片中没有其他对象的前提下，蓝色小汽车怎么也不会和相关列表上跑到红色小汽车或含有红色的小汽车的前面去。在图例查询中，系统的目标已经从给出确定性的答案转变为尽可能地减少用户搜索和评价工作量并提高用户的满意度。

图例查询向研究人员提出了新的挑战。除了选择适当的特征外，还必须定义出有效的相似匹配算子。与精确匹配算子不同，特征的相似匹配算子可以有多种定义方法和途径。文献[14]给出了对一个特定特征的两个值计算其距离的六种不同方法，每个距离算子在检索性能上表现出了很大的差异。

除了图像特征的选取和相似度的评价外，用来查询数据库的示例图像也是十分重要的。下面我们根据示例图像的来源不同将图例查询分为三种类型分别加以讨论。

2.1.2.1 外部图例查询 (Query by External Pictorial Example)

外部图例查询中的示例图像来源于检索系统的外部，也就是说给定的图像不是数据库中存储的图像。用户可以数字化一张照片，也可以在 Internet 或别的什么地方找一幅质量比较好的图片作为示例，要求系统在数据库中搜索相似的图像。

例如，当一个企业需要注册商标的时候，可以将其新的商标作为示例检索数据库，以确定没有太多相似的已注册的商标。执法机构也可以使用外部图例查询将现场犯罪痕迹与犯罪记录库的信息相比较寻找破案线索。已经有系统成功地根据一个指纹在指纹数据库中快速地确定指纹所有者，而且数据库可能十分庞大。1994 年美国联邦调查局 (FBI) 的指纹档案中存储了 1 亿多张指纹卡片，而每张卡片又包含了 10 个指纹。我们还可以举出一些外部图例在执法过程中的应用，比如警察可以向检索系统提供保安摄像机拍摄的劫匪的照片，系统在警察局存档的嫌疑犯照片库中搜索并返回与之相似的图像列表。

外部图例查询的优势在于用户在表达其信息需求时的简单性。用户仅需提供一幅或多幅图像，或许也需要调整一些参数，如各种特征的权重等。检索过程中并不需要特定的数据库查询语言方面的知识。

我们同时应该注意的是，尽管用户向系统提供了自己认为很重要或很有代表性的图像，系统也有可能不能精确地满足用户所希望的相似性需要。“相似性”对系统而言是一些特征值的相似，虽然系统尽可能地用特征值相似来模拟人类视觉甚至感觉上的相似性，但这两者之间总不可避免地存在一些差异。这一点常常使得那些使用外部图例查询的用户中的新手感到困惑不解。

当用户很难提供一个示例图像时，外部图例查询的可用性将大打折扣。也许在某种情况下，在系统外部寻找一幅示例图片所付出的代价甚至有可能超过直接在数据库中寻找图像的代价。

总之，外部图例查询的优点是易用性，缺点是需要付出一定的努力向系统提供外部示例图像。

2.1.2.2 内部图例查询 (Query by Internal Pictorial Example)

内部图例查询中的示例图片是从检索系统的数据库中选出的。当没有可用的外部示例图片时，用户可以从可用的图像集合中选择查询图片。系统在其他方面的功能与外部示例图片基本上一致。

为了比较内部与外部示例查询，我们拿前面提到的抢劫嫌疑犯的照片的检索作例子。如果保安摄像机没有拍摄到有效的劫匪的照片，警察只有依赖于目击者。目击者没有必要一张一张地翻阅警察局存档的嫌疑犯照片库，通过图例查询，目击者可以先选择一张看上去比较象嫌疑犯的照片，比如两者都有络腮胡子，然后从系统给出的所有相似者列表中快速确定搜索目标。

内部图例查询的优势在于不再需要为了进行图例查询而首先从外界寻求或构造一幅示例图片。但是，系统要求用户必须首先在数据库中找到一幅类似的图片，这个过程可能是相当费时费力的。所以内部图例查询经常与其它检索方法结合使用以避免这个问题。比如可以先通过直接查询缩小图像搜索范围，然后再使用内部图例查询。

总之，内部图例查询的优点是用户不再为获得一个外部示例图片而烦恼，缺点是仍然需要用户付出努力在数据库中找到一幅合适的示例图片。

2.1.2.3 草图查询(Query by Sketch)

草图查询是图例查询的一种形式，其中的示例图像由用户创建。用户为他希望的得到的图像勾画一张草图，所以也有人把这种查询机制称之为查询画布 (Query Canvas) (文献 [13])。用户在查询画布上组织或编辑一些预定义的图片元素以创建一幅示例图像。检索系统为此必须为用户提供一些图像部件 (如各类纹理) 和作图工具。

我们还是拿前面提到的抢劫嫌疑犯的照片检索作例子，这里向系统输入的示例图像是一张嫌疑犯的简略肖像草图，它是由警察中的艺术家根据目击者的叙述描画的。有时目击者可以通过组合一些诸如头发、鼻子、眼睛等部件的照片从而构造一张脸。

与自然图像相比，人工创作的草图具有更高的抽象性，用户只需勾画出他所寻找的图片的重要部分，在这方面草图查询要比其他使用已有图片的图例查询更具优势。

草图查询的缺点是它要求用户或多或少要有一些艺术才能。由于大多数的用户不具备足够的这方面的才能，草图查询在大多数的场合下仅用来指出图像中对象的位置或一些对象的全局属性。

可见，草图查询的优点优点是用户可以根据需要指出最重要、最感兴趣的图像细节，缺点是草图的创建不象想象的那么容易，而且很难建立草图与图像之间的映射关系。

2.1.3 浏览(Browsing)

前面所讨论的这些检索方法都被看成是如何形成一个有效的数据库查询的过程。这些

方法使用不同的抽象类型,在查询表达式和用户输入方面有互不相同的形式。但按照文献[13]的说法,所有这些方法都使用同一种数据导航方式,这些方法的目标都是要定位信息。在定位导航方式下,用户应该首先定义其信息需求,然后检索系统根据这个需求进行搜索。文献[13]指出的另外一种不同的数据导航方式为浏览,在这种方式下,用户并不是为了寻找特定的,或者说是明确的图像。

尽管“浏览”这个词在各类文献中被频繁使用,但很少有人为它给出一个明确的、正式的定义,这是因为浏览具有直觉和非正式的特性。由于缺乏清晰的、广泛接受的定义,“浏览”现在被用来表示多种信息检索方式。本文中的“浏览”有其专门的含义:浏览是一种由人进行的特别的信息搜索。

尽管很难给出浏览的定义,但在各类文献中对这种信息检索方式的特性都具有一致的观点。在这里我们给出浏览的三个主要的特性:

1) 这种信息检索方式具有很强的交互性,其中用户的角色是非常重要的。它借助于人类的认知能力来避免人机交互的某些问题,它的依据是人类在别期望的信息比描述其信息需求更具能力上的优势。

2) 用户的搜索路径是不确定的,数据库中的记录被评价的顺序完全取决于用户的选择。浏览是一种特殊的、探测性的过程,用户连续地改变其搜索方向以及对自己信息需求的感知,比如,信息需求可以从“一幅好看的图像”的初始需求逐渐演变为“华山日出”的特定搜索。同时,不同的用户在实现同样的信息需求时可能通过不同的途径。文献[18]把浏览的这种特性称之为“主题的偶然组配”(Serendipity),而 Serendipity 的本意是指不经意发现新奇事物的天赋,这里是指人所具有的揭示数据库记录及记录之间信息和联系的天赋。

3) 用户向系统提出其数据或信息要求并不需要用户具有查询语言或数据库体系结构的知识。比如在一种简单的浏览方式下,用户只需点击“下一页”或“更多的图片”按钮来表达其对新的数据集的需求。浏览的这种用户友好性使得浏览检索的成功很大程度上依赖于用户的行为。

当用户不能明确定义其信息需求,或者用户只是想获得对数据集的印象时,浏览是一种比较适合的检索方式。在前一种情况下,浏览可以辅助用户确定其信息需求;在后一种情况下,浏览可以通过顺序地或随机地逐步提供所有可用数据使得用户获得对整个数据集的整体观点。基于浏览的这两个功能,我们将图像检索领域中的浏览区分为下面两种形式加以讨论

2.1.3.1 直观浏览 (Visual Browsing)

直观浏览是一种直接由用户一幅一幅地观察图像的检索方式,是一种最简单的图像检索形式。

这种方法的优点是,用户在浏览过程中要么看到了他所寻找的图片,要么发觉图片不在当前集合中。用户不必去构造一个查询表达式,甚至无需就其需求与系统进行沟通。

直观浏览的缺陷是,用户在发现其需要的图片并不在图像集合中之前不得不观察所有的图片。同时,即便最终发现了需要的图片,在此之前也必须察看许多的图片。平均意义上来说,用户在发现一张需要的特定的图片之前将遍历一半的图像集合。

由此可见,直观浏览特别适用于对小规模的图像集合进行快速的扫描。

2.1.3.2 联接查询(Query by Association)

顾名思义,联接查询是一种较之直观浏览更为结构化的浏览形式,它通过对检索出的数据进行选择来检索与之相联接的数据。

由于超文本系统的引入,联接查询已经一种十分普遍的检索方式。目前,超文本是检索文本信息的一种重要方法,当除了文本以外由其他媒体引入数据库时,我们常使用“超媒

体”一词。

超媒体众所周知的应用有 Windows™的帮助系统和万维网 (World Wide Web), 在这些系统中, 联接建立在两个点之间 (有时我们称之为 Anchor), 比如从一个词组到一幅图像的联接, 当用户选中这个词组时, 它将跳转到词组所指向的图像。尽管大多数情况下联接是预定义的, 但联接不一定就必须是静态的。如果源 Anchor 与一个查询相联接, 源 Anchor 将指向查询的结果, 这方面一个很好的例子就是 WWW 上的动态页面 (Dynamic Pages)。

菜单机构是一种层次结构的联接查询。通过选择菜单项, 用户可以逐步细化其查询。例如, 用户从商品目录中首先选择“交通工具”, 然后从“汽车”、“火车”、“飞机”和“轮船”中选择“汽车”, 最后从一系列的汽车牌号中选了他所需要的“桑塔纳 2000”, 并检索所有有关“桑塔纳 2000”的图片。在这个过程中系统使用了预定义的结构对数据库中的图像进行分组。

联接查询的优点是, 用户可以在图像预定义的语义导航下在数据库中检索图像。这时“查询”工作可能简单到用户只需在数据选项 (可能是图像、图像的描述或图像的抽象特征等) 上点击鼠标, 系统随后给出与之相联接的所有数据。

类似于根据描述的直接查询, 联接查询在语义定义上是相当困难的, 需要相当大的努力对每一幅图像进行图像分析, 而且通常由人工完成。

由此可见, 联接查询一般适用于图像可根据其内容进行分组的应用。

2.1.4 综合检索方法

大多数的图像检索系统都或多或少地综合应用了上述的各种检索方法。系统可以对同一个数据库提供多种检索方法的接口, 但更有意义的是在一次检索会话中综合应用不同的检索技术。后一种情况下, 检索系统将是多个应用不同检索方法的子系统的有机集成, 子系统之间通过输入与输出的对接完成通讯与交互。

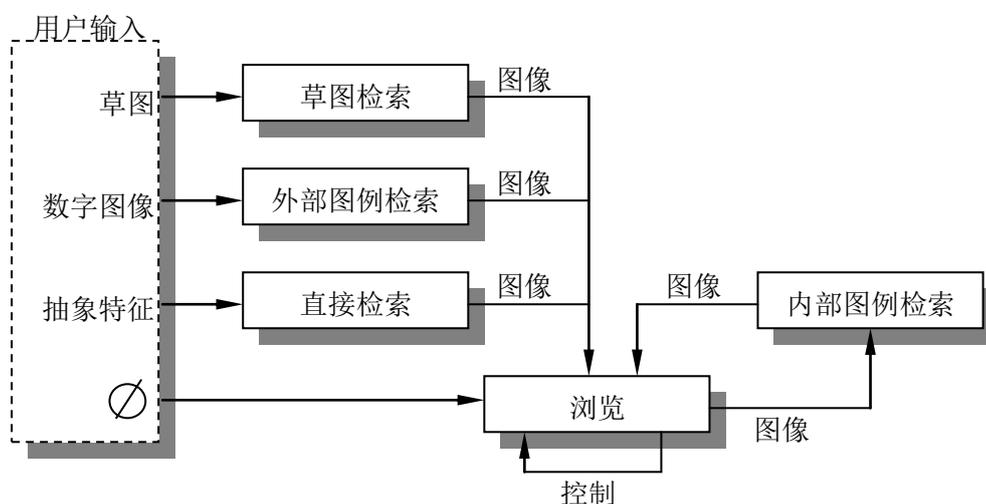


图 2-2 综合运用各种检索方法

能够进行集成的各种检索方法的子系统的数目常常受到子系统之间交互能力的限制。图像检索系统流程图 (图 2-2) 描述了一种检索方法进行集成的可能性。“用户输入”子系统主要考虑到用户输入的几种可能。如果用户向系统输入示例图像, 只有外部图例查询和草图查询是可用的; 如果用户输入包含抽象特征, 系统将提供直接查询服务; 当然用户也有可能不给系统提供任何输入, 但希望系统给出有关数据库内容的概要信息, 这时将由系统而不是用户直接初始化检索过程, 用户将直接浏览数据库内容。

所有交互式检索方法将检索出的图像显示给用户, 这也正是这类系统的目的。用户将

检阅输出的图像结果集,此时用户的行为即为直观浏览,几乎所有的图像检索系统都不可避免地要使用这一检索方法。当然,极端的情况下,某些检索功能较弱的系统有可能只提供了直观浏览的检索方法。

外部图例查询、草图查询和两种直接查询的输出将送至直观浏览显示给用户,这四种方法并不使用浏览检索的输出结果,而这个结果是由于浏览检索过程中用户做出选择而产生的。草图查询需要有用户创作的输入图像,而外部图例查询需要从数据库以外提供一个示例图像,如果示例图像来源于数据库,系统可使用内部图例查询手段提供检索服务。

从图 2-2 中可以看出,所有三大类检索系统的输出最终将都汇集到浏览检索,而浏览检索方法的输出有两种输出形式:

1) 用户选择的图像。如前所述,能够将数据库中的图像作为输入的唯一检索方法是内部图例查询。所以,在浏览之后,内部图例查询成为除了浏览之外唯一可用的检索方法。

2) 用户控制。在直观浏览的方式下,用户控制可以辅助图像的显示。通常,用户需要诸如“第一组”、“最后一组”、“下一组”、“前一组”等的导航按钮来辅助浏览。在联接查询的情况下,控制担任着更加复杂、更加细致的任务,这时控制意味着用户希望看到与当前选项相联系的图像。如果选项是一幅图像,则有两种可能,一种是将图像本身作为内部图例查询的输入进行检索,另一种可能是图像是用来指出其符号含义,该含义将被用来进行联接检索,究竟是哪一种情况完全取决于系统对该图像作用的定义。

图 2-2 不仅给出了在一个系统中多种检索方法集成的可能性,而且还描述了系统内的交互。外部图例查询、草图查询和两种直接查询接受用户的输入并将输出送至直观浏览,如果用户对输出不满意,则需要启动一个新的检索会话。用户也可以利用另外的三个检索方法逐步细化它的查询,当然直观浏览的作用是有限的,而且工作方式相当琐碎。内部图例查询和联接查询将使得用户可以对检索结果进行迭代求精甚至重新构造信息需求,这两种方法在交互性能和迭代检索性能上优于其它的检索方式。

总之,图像检索系统在集成各类检索方法时应考虑到不同的方法的功能和限制,使其在不同的应用场合和检索过程中发挥出最大优势。

§ 2.2 图像检索模型

为了能够有效地确定一个图像检索系统的特点,比较和分析各类图像检索系统,这里我们将给出一个图像检索模型。这个模型是以 Infoscope[13]的高层结构和 Grosky[12]多媒体信息系统的通用结构为基础的。这两种结构都是一种通用的多媒体信息系统结构,但是 Infoscope 更侧重于图像数据的处理。首先我们对这两种结构作一简单的介绍。

2.2.1 Infoscope

Infoscope 是一种信息检索系统。Infoscope 的高层结构主要由 4 个主要部件构成:

- (1) 数据库模块存储数据和所有导出特征;
- (2) 插入模块从图像中抽取特征并将其(图像及其特征)插入到数据库中;
- (3) 接口模块提供人机交互。人机接口辅助用户形式化其信息需求并负责向用户提交系统检索结果;

(4) 知识库中包含了计算和比较特征值的方法以及应用的领域知识。文献[13]还专门就在图像信息系统中引入领域知识的重要性进行了论述。

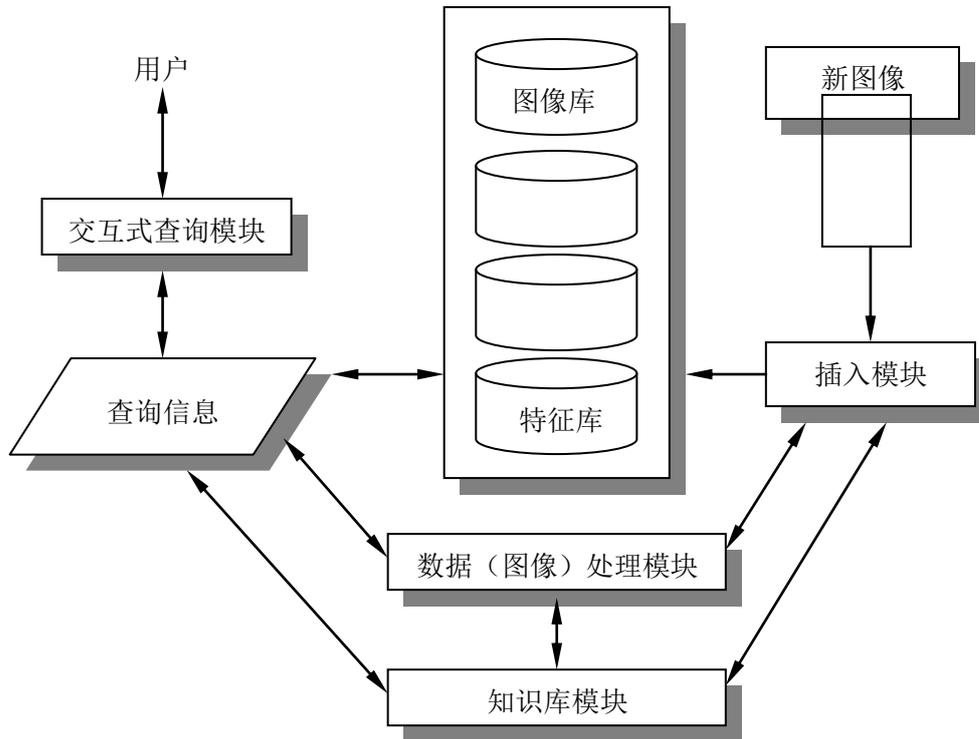


图 2-3 Infoscope 系统结构

图 2-3 给出了 Infoscope 的系统结构，其中还包含了除上述四种主要部件以外的其他一些模块。交互式查询模块和查询形式化模块构成了接口模块，图像处理模块方便了根据知识库信息进行的解释和描述。被处理的图像主要有两个来源，一个是为了进行图像数据采集由插入模块给出的图像，另一个是为了解释示例图像而由接口模块给出的图像。

2.2.2 Grosky

文献[12]详细论述了 Grosky 多媒体信息系统的结构（如图 2-4 所示）。其中数据库部件主要由三个逻辑上独立的库（Repository）组成：

- (1) 标准的文字和数字库，用来存储客观世界中实体的属性（Facts）；
- (2) 多媒体数据库，用来存储未解释的多媒体对象，在本文中主要指的是原始图像数据。
- (3) 特征库，用来存放通过特征处理模块从多媒体对象中抽取的特征值。

其中的插入模块的作用仅限于将多媒体对象插入到数据库中，而在 Infoscope 中的插入模块完成的图像解释与抽象工作则交给了特征处理模块。合成模块的作用是将多个多媒体对象进行组合合成。在整个结构中没有提供从系统到用户的信息流。

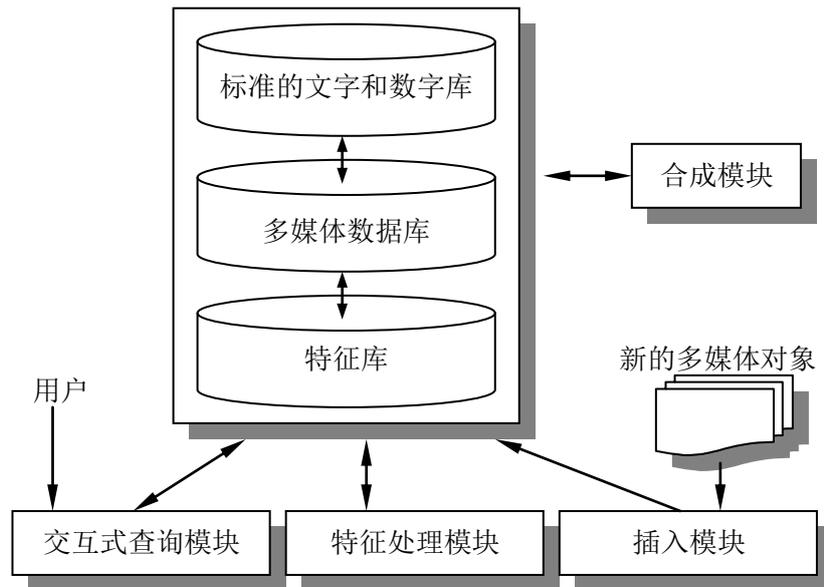


图 2-4 Grosky 多媒体信息系统的结构

2.2.3 图像检索模型

这里将给出一个用于图像数据库系统的图像检索模型（Image Retrieval Model），这个模型不仅可以用来作为描述和比较图像检索方法的一般框架，也可以作为实现图像检索机制的参考结构。

为了讨论方便起见，我们首先做出两个假定：

1. 我们假设图像数据库中所有的图像是一次输入的，也就是说一旦数据库投入运行（这里仅指检索操作）后，就不再有图像插入。当然，这个模型稍作扩充即可处理动态模型，只是为了简单地进行数据流的描述而将其简化。
2. 数据库初建时除了允许输入原始图像数据外，其他多媒体方面的需求已经进行了必要的限制，使我们能集中精力考虑与图像有关的问题。

一个图像检索模型主要由五个部件组成：

1. 图像分析模块（Image Analysis Module）。将原始的数字图像作为其输入，并从中抽取所需的信息。分析过程可以由人来完成，也可以由计算机自动地或半自动地进行。
2. 数据库模块（Database Module）。其中存放了所有的数据。在数据库的采集过程中，图像分析模块将图像及其相关信息提交到数据库。数据库模块由三个逻辑上的仓库（Repository）组成：
 - 图像库（Image Repository），存储了所有的原始图像数据，可以是整幅图像或图像的一部分，如图像中某个用以标示特定对象的区域描述。
 - 特征库（Features Repository），由它来掌握图像库中所有数据的特征抽象信息，如关键字，度量信息等。包括基于内容的特征，也包括基于文字的特征。
 - 联接库（Link Repository），包括图像库中数据记录之间准确定义的联系信息，如一辆小汽车的照片和小汽车主人的照片之间的关系，某个病人头部多张 CT 切片之间的位置与距离关系等。

这样，从图像分析模块提交的有关的图像信息可以分为两种类型，描述图像一系列特征的信息将被存入特征库，而图像与图像之间关系信息将被存入联接库。数据库模块同时应支持数据的一般操纵，如插入、删除、修改，并保证数据的完整性与一致性。一般情况下，我们可以用一个支持多媒体数据存取的完备的关系数据库

来实现数据库模块。

3. 用户模块 (User Model)。用以提供系统必需的或希望得到的某些信息。由用户来决定检索过程的开始和结束。这里，“用户”可能是直接使用检索系统的人，也可能是为另外一个信息系统提供图像检索服务的接口模块。
4. 交互查询模块 (Interactive Query Model)。用以将用户的需求信息映射成为数据库的查询并将查询的结果提交给用户。比如，当用户将其需求信息以图像的形式提交给系统后，交互查询模块将要求图像分析模块对这个图像进行必要的解释，然后根据分析的结果将用户的需求信息用数据库支持的语言（如 SQL）表述的查询并提交给数据库模块，数据库模块根据查询要求完成实际数据库检索工作，最后由交互查询模块将结果以适当的形式返回给用户。与此同时，交互查询模块应根据用户对查询结果的意见进一步改进查询，以针对某个特定用户迅速掌握其查询的意图和倾向。交互查询模块可以利用知识库模块提高命中率。
5. 知识库模块 (Knowledge Model)。知识库中应包括特征的使用信息和图像相似比较的评价模型，在用户的需求信息和数据库提供的信息（如统计信息）的基础上，知识库模块应为交互查询模块提供复杂有效的查询引导。

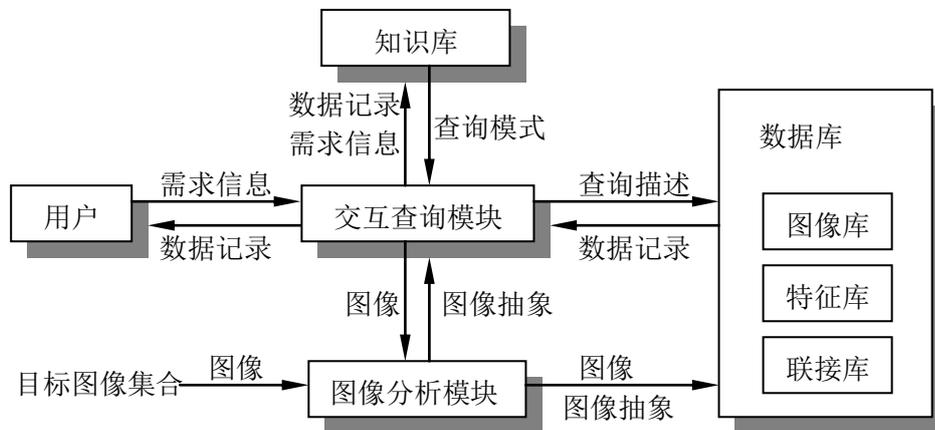


图 2-5 图像检索模型

图 2-5 所示的图像检索模型的部件是与模块间的主要数据流一起描述的。这些部件支持图像检索系统赖以工作的两个主要处理过程的执行：

1. 图像采集过程，有时我们也称之为插入或数据捕获，用来将图像数据插入到特定的数据库中。
2. 图像检索过程，用来寻找和提供存储在数据库中的特定图像。

数据库采集处理只用到了图像分析模块和数据库模块，而检索处理可能需要所有五个部件的支持。

在这个模型中，图像分析模块与 Grosky 中的特征处理模块以及 Infoscope 中的图像处理模块在功能上是一致的。但应注意到图像理解方面的知识已经包含在图像分析模块中了，而 Infoscope 的图像处理模块必须与独立的知识模块交互以获得这方面的知识。图像分析模块实际上综合了 Infoscope 中的插入模块、数据（图像）处理模块和一部分有关图像处理方面的知识。将这三个模块综合在一起看待问题能够减少模块间的耦合复杂度，而且可以使得我们能够集中精力研究主要系统部件间的数据流。

与 Grosky 不同的是，我们提出的图像检索模型只允许图像分析模块向数据库输入数据，而且特征库综合了 Grosky 中的标准的文字数字库和特征库，Grosky 之所以要把这两个库单独出来分别考虑，是因为它需要处理多种形式的多媒体数据。联接库是对 Infoscope 和 Grosky 中的数据库模块的扩充。

用户提出的对一幅图像进行解释的需求将直接交给图像分析模块处理，而 Infoscope 却令人费解地要通过数据库模块的支持。

由此可见，本文所提出的图像检索模型比 Infoscope 的模型更加简洁，也比 Grosky 的模型更加清晰，而且，不被由于考虑图像之外的其他多媒体对象的检索而导致的特殊应用环境所干扰，所以本文提出的模型在描述与比较图像检索方法时作为一般的框架将更为有效。进一步的详细内容可以参阅有关文献。

本文提出的图像检索模型不仅可以用来作为描述和比较图像检索方法的一般框架，也可以作为实现图像检索系统的指导性结构，具有较强的实用性。作为应用实例，也为了评价现有的各种图像检索手段，下面我们将应用这个模型对多种图像检索方法进行分析与比较。

§ 2.3 用图像检索模型分析检索方法

这里我们将用图像检索模型的观点描述并分析这些方法。对每种图像检索方法，我们将考查数据库的内容、图像采集过程和检索过程，而把用户模块和目标图像集合视为常量。数据库模块总包含一个非空的图像库，特征库和联接库如不另作说明则总认为是空的。

2.3.1 直接查询 (Direct Query)

2.3.1.1 根据描述查询

在图像采集过程中，由图像分析模块对每幅图像定义一些关键字，并插入数据库。这个过程可以是自动的或半自动的，通常基于其他媒体类型或属性数据进行。在实际应用中，大多数的系统将人作为图像分析“模块”来完成图像评价和关键字定义。

在检索过程中，用户模块给系统输入关键字来表述其信息需求。交互查询模块可以辅助用户描述其信息需求，比如给出数据库中某关键字可用信息的列表供用户选择。交互查询模块根据给定的关键字查询数据库，如

```
SELECT images FROM database WHERE keyword = 'CAR';
```

这里知识库模块不是必须的，但可以用来帮助提高检索性能。首先它可以提供基本的全文本检索技术，比如辞典技术。另外，比如当用户提出检索“运输”图像时，知识库模块也可以将其变换为对一些诸如‘CAR’、‘BIKE’、‘PLANE’等关键字的检索。

这里，特征库中只需包含文字与数字数据（关键字）。

2.3.1.2 根据图像特征查询

在图像采集过程中，由图像分析模块对每幅图像计算其一系列的度量信息（如大小，颜色数，平均能量等），并插入数据库。这个过程可以完全自动化。

在图像检索时，用户可以针对检索目标给出某些度量的值或值的范围，由交互查询模块直接将用户查询映射为数据库查询以检索满足条件的图像。

在数据库中，特征库包含从图像得到的度量信息。

2.3.2 图例查询 (Query By Pictorial Example)

对下述的三种图例查询而言，数据库的采集过程是一样的。在数据库的采集过程中，图像分析模块计算图像的一些度量信息，然后图像及其度量存入数据库。和根据特征的直接查询相似，特征的抽取是完全自动化或半自动化的。

同样，三种图例查询中，数据库模块的特征库都将用来存储图像的度量信息。

2.3.2.1 外部图例查询

检索过程中的用户输入由一个或多个数字图像组成，并对相似度的度量在不同的方向或分量上根据其重要程度由用户赋予不同的权重。交互查询模块从图像分析模块得到用户图片的度量信息，并据此将用户的查询映射到数据库查询。知识库模块可以根据用户提供的权重提供相似度的分级方式。

2.3.2.2 内部图例查询

用户模块的首次输入是以交互查询模块提供的信息为基础的。交互查询模块向知识库模块请求将提供的图像集合进行分类合成，通常知识库模块随机地或根据一定的预定义规则选取图像。用户从选出的图像中再选择一幅或多幅图像，用户的选择将被交互查询模块转换为数据库查询。

跟外部图例查询一样，用户可以通过定义权重来标明不同特征的不同重要程度，知识库模块提供相似度分级方式。除了用户提供的权重外，知识库模块在组织查询模式时可以同时考虑用户先前的选择。对交互查询模块作为查询结果提供的图像集合，用户可以做出新一轮的选择，除非用户决定中止这个过程，这种交互式的检索将继续下去。

2.3.2.3 用户草图查询

根据用户草图查询的检索过程与外部图例查询非常相似。取代完整数字图像的是，用户模块的输入是对要寻找的目标图像的印象，用户可以用手绘草图来描述其印象，当然也有可能对在草图中使用到的特征赋予不同的权重。交互查询模块应辅助用户完成草图的绘制。图像分析模块将解释草图，交互查询模块将形成数据库查询。根据用户提供的权重，知识库模块将为此建立适当的分级方式。

2.3.3 浏览 (Browsing)

对下述两种浏览方法而言，最显著的特征是在检索过程的开始时，由交互查询模块初始检索过程，用户再对系统提供的信息做出反应。我们将看到，用图像检索模型来考察直观浏览 (Visual Browsing) 和联接查询 (Query By Association) 时，将会发现它们有很多的不同之处。

2.3.3.1 直观浏览

直观浏览是一种最简单图像检索方式。图像采集过程将图像集合输入到数据库中。图像分析模块的作用仅是传递图像，而不解释图像，在直观浏览的方式下该模块是一个哑元 (dummy) 模块。

在纯粹的直观浏览方式下，交互查询模块的作用仅是在数据库中选择所有的图像并显示给用户。实践中，直观浏览常在图像集合的表达或显示时增加一些控制，如“下 20 个图像”、“下一屏”、“前一幅”等。直观浏览不使用知识库模块。

2.3.3.2 联接查询

在其他检索方式的数据库采集过程中，图像是单独地进行分析的。在联接查询中，图像分析模块的分析解释工作只是将当前图像与另外一些图像相联系。图像分析模块将图像集合中图像之间的联系 (上文中我们称之为“联接”) 送至数据库模块。在联接是动态的情况下，联接是根据图像抽象的规则定义的，而不是硬性定义的。比如，一个静态联接可能是“图像 X 与图像 Y 具有某种联系”，而一个动态联接则可能是一个规则，如“所有具有关键字‘CAR’的图像之间相互联系”。

交互查询模块在知识库模块提供的查询模式基础上，向用户提供数据库中的记录 (图

像或特征抽象), 从而启动检索过程, 查询模式使得为一次浏览会话从联接库中选择一个适当的起点成为可能。随后, 用户选择其中的一个记录以便检索所有与之联接的记录。

在数据库中, 联接库存储了所有图像之间的联系。特征库用来存放从图像抽象出的特征, 但此时特征库不是必须的。

§ 2.4 小结

本章工作的主要背景是我们进行的医学影像数据库的研究中有关检索系统的通用结构的一些成果。我们使用本章提出的图像检索模型分析比较了现有的图像检索技术, 结合医学影像的特点, 建立了医学影像的原型数据库系统, 实践证明该模型在检索技术分析、检索系统构造上的指导和参考作用是清晰有效的。在下一章中我们将给出医学影像数据库系统的具体结构设计。

正如我们在本文开始时指出的那样, 为了简化叙述与理解, 我们提出了两个假定。在实际应用中, 应考虑到图像数据库的动态特性以及具体应用环境的需求, 如专业领域知识在检索中的导引作用、当存在其他多媒体信息时多种信息表达方式之间的联系与同步等。我们可以预见, 在本文提出的模型基础上针对不同的应用做出一定的扩展是方便的。

第三章 医学影像数据库系统的设计

在上一章中我们给出了一个图像检索模型，我们将以这个模型为基础来进行医学影像数据库的设计。从图 2-5 中可以看出，图像检索模型中各个模块功能相对独立，模块之间具有典型的层次结构。我们看到用户模块是用户交互和系统功能表示的接口，图像分析模块、知识库模块和交互查询模块形成了系统功能的应用逻辑，而数据库模块则提供数据服务。这样我们看到了一种典型的三层结构模型。

三层 Client/Server 的数据库系统结构能够很好地实现这种图像检索模型。同时医学影像数据库具有普通数据库所具有的一切需要和特征，其更加重要的一个目的是医学影像资料的管理。在下面我们将看到在其它数据服务方面，三层 Client/Server 结构的优势也是明显的。

这样，医学影像数据库系统的设计将是医学影像管理需求和图像检索模型在三层 Client/Server 结构上的贯彻与落实。

§ 3.1 三层 Client/Server 结构

Client/Server 结构是一种两层结构的模式，在客户端集中了图形用户界面（GUI）和应用程序模块，数据存取模块则放在服务器端，这种结构易于开发复杂的应用，但是依赖于客户端的能力，造成客户端软件负担重，即所谓的胖客户（Fat Client），难以适应大规模的系统需要。从可移植性而言，用于开发 Client/Server 客户端的开发工具是专业化的，不同工具的代码是互不兼容的，从长远的观点看，就使得两层结构系统的实现存在冲突，这意味着将来一种专用技术向另一种专用技术的转移，也就带来对原有投资的浪费。另外，由于 Client/Server 结构没有设置中间层来提高安全性，所以系统的安全保密性欠佳。

而三层信息系统结构通常指在 Client/Server 系统中，把应用系统分为三个部分：实现用户接口部分的构成表达层，对数据进行处理的成功能层，第三部分是用来放置数据基的数据层。

在移植性方面，三层结构中按标准化设计的组件可以被重复使用，功能性的中间层可以被不同平台的不同客户工具访问。并且在中间层的服务器中提供的加密核心技术使得这种结构的安全保密性能大大增强。另外，由于是按层分割功能，改变任一层的实现逻辑都不会影响到其他层，利于变更和维护应用技术规范，而且使得表达层的图形用户界面的设计灵活性很大。在开发速度上，虽然初次开发时，两层结构开发速度快，但在二次开发时，三层结构比两层结构速度快得多，如果有可重用库和中间层逻辑的话，二次开发的时间将进一步缩短。

3.1.1 三层 Client/Server 的基本硬件结构

传统的二层 Client/Server 结构存在以下几个局限：

- （1）它是单一服务器且以局域网为中心的，所以难以扩展至大型广域网或 Internet；
- （2）受限于供应商；
- （3）软、硬件的组合及集成能力有限；
- （4）难以管理大量的客户机。

因此，三层 Client/Server 结构应运而生。三层 Client/Server 结构是将应用功能分成表示层、功能层和数据层三部分。其解决方案是：对这三层进行明确分割，并在逻辑上使其独立。原来的数据层作为 DBMS 已经独立出来，所以关键是要将表示层和功能层分离成各自独立

的程序，并且还要使这两层间的接口简洁明了。

将上述三层功能装载到硬件的方法基本上有三种（如图 3-1 所示）。其中表示层配置在客户机中，而数据层配置在服务器中。一般情况是只将表示层配置在客户机中，如图 3-1 (a) 或图 3-1(b)所示。如果像图 3-1(c)所示的那样连功能层也放在客户机中，与二层 Client/Server 结构相比，其程序的可维护性要好得多，但是其他问题并未得到解决。客户机的负荷太重，其业务处理所需的数据要从服务器传给客户机，所以系统的性能容易变坏。

如果将功能层和数据层分别放在不同的服务器中（如图 3-1 (b) 所示），则服务器和服务器之间也要进行数据传送。但是，由于在这种形态中三层是分别放在各自不同的硬件系统上的，所以灵活性很高，能够适应客户机数目的增加和处理负荷的变动。例如，在追加新业务处理时，可以相应增加装载功能层的服务器。因此，系统规模越大这种形态的优点就越显著。

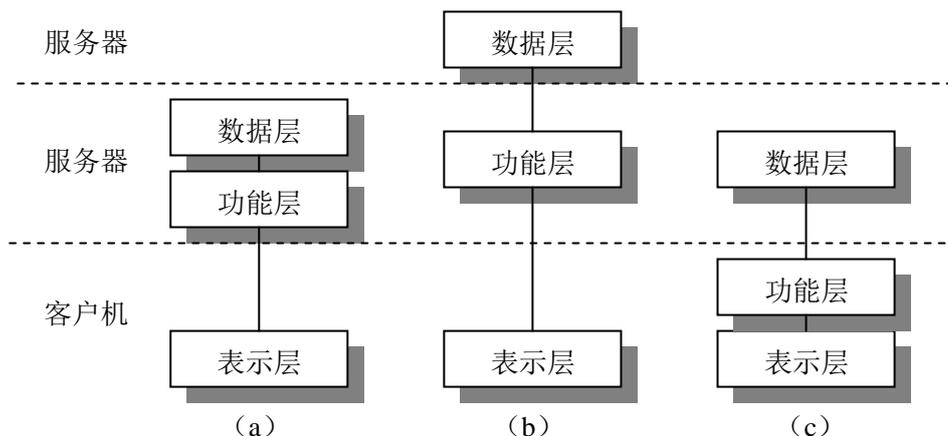


图 3-1 三层 Client/Server 的基本硬件结构

值得注意的是：三层 Client/Server 结构各层间的通信效率若不高，即使分配给各层的硬件能力很强，其作为整体来说也达不到所要求的性能。所以，设计时必须慎重考虑三层间的通信方法、通信频度及数据量。这和提高各层的独立性一样是三层 Client/Server 结构的关键问题。

在医学影像数据库的设计中，由于需要在各层之间传送大量的图像资料，所以通讯问题尤为突出。由于在物理实现上，可能存在着多种通讯方式，可能是通讯速率较高的局域网，也可能是通讯速率很低的 Internet（在基于 Web 的实现时），所以在进行系统设计时，要更多地从软件设计的角度充分考虑到这一点。本章的最后将就此问题专门讨论并将提出一些对策。

3.1.2 三层 Client/Server 的功能

3.1.2.1 表示层

表示层是应用的用户接口部分，它担负着用户与应用间的对话功能。它用于检查用户从键盘等输入的数据，显示应用输出的数据。为使用户能直观地进行操作，一般要使用图形用户接口(GUI)，操作简单、易学易用。在变更用户接口时，只需改写显示控制和数据检查程序，而不影响其他两层。检查的内容也只限于数据的形式和值的范围，不包括有关业务本身的处理逻辑。

图形界面的结构是不固定的，这便于以后能灵活地进行变更。例如，在一个窗口中不是放入几个功能，而是按功能分割窗口，以便使每个窗口的功能简洁单纯。在这层的程序开发中主要是使用可视化编程工具。

表示层的另一种实现方式是基于 Web 浏览器的实现，可以通过浏览器的表单、JavaApplet 的 CUI、ActiveX 等技术或方式实现表示层逻辑。而在功能层和数据层上则通过 Web 服务器、中间件、数据库服务实现，这就形成了所谓的三层 Web/Server 模式。从本质上来讲，Web/Server 是三层 Client/Server 结构的一种。

3.1.2.2 功能层

功能层相当于应用的主体，它是将具体的业务处理逻辑地编入程序中，而处理所需的数据则要从表示层或数据层取得。

表示层和功能层之间的数据交往要尽可能简洁。例如，用户检索数据时，要设法将有关检索要求的信息一次传送给功能层，而由功能层处理过的检索结果数据也一次传送给表示层。在应用设计中，应尽量避免进行一次业务处理，在表示层和功能层间进行多几次数据交换的笨拙设计。

通常，在功能层中还包含有确认用户对应用和数据库存取权限的功能以及记录系统处理日志的功能。这层的程序多半是用可视化编程工具开发的，也有使用 COBOL 和 C 语言的。

3.1.2.3 数据层

数据层就是 DBMS，负责管理对数据库数据的读写。

DBMS 必须能迅速执行大量数据的更新和检索。现在的主流是关系数据库管理系统 (RDBMS)。因此，一般从功能层传送到数据层的要求大都使用 SQL 语言。

3.1.3 三层 Client/Server 结构的优点

(1) 具有灵活的硬件系统构成。对于各个层可以选择与其处理负荷和处理特性相适应的硬件。这是一个与系统可缩放性直接相关的问题。例如，最初用一台 Unix 工作站作为服务器，将数据层和功能层都配置在这台服务器上。随着业务的发展，用户数和数据量逐渐增加，这时就可以将 Unix 工作站作为功能层的专用服务器，另外追加一台专用于数据层的服务器。若业务进一步扩大，用户数进一步增加，则可以继续增加功能层的服务器数目，用以分割数据业务。清晰、合理地分割三层结构并使其独立，可以使系统构成的变更非常简单，被分成三层的应用基本上不需要修正。

(2) 提高程序的可维护性。三层 Client/Server 结构中，应用的各层可以并行开发，各层也可以选择各自最适合的开发语言。

(3) 利于变更和维护应用技术规范。因为是按层分割功能，所以各个程序的处理逻辑变得十分简单。

(4) 进行严密的安全管理。越关键的应用，用户的识别和存取权限设定愈重要。在三层 Client/Server 结构中，识别用户的机构是按层来构筑的，对应用和数据的存取权限也可以按层进行设定。例如，即使外部的入侵者突破了表示层的安全防线，若在功能层中备有另外的安全机构，系统也可以阻止入侵者进入其他部分。

此外，三层 Client/Server 结构还具有系统管理简单，可支持异种数据库，有很高的可用性等优点。

3.1.4 中间件

每个 Client/Server 环境，从最小的 LAN 环境到超级网络环境，都使用某种形式的中间件。实际上，无论客户机何时给服务器发送请求，也无论它何时应用存取数据库文件，都有某种形式的中间件传递 Client/Server 链路，用以消除通信协议、数据库查询语言、应用逻辑

与操作系统之间潜在的不兼容问题。中间件是 Client/Server 环境中最重要的部件。

所谓中间件是一个用 API 定义的软件层，是具有强大通信能力和良好可扩展性的分布式软件管理框架。它的功能是在客户机和服务器或者服务器和服务器之间传送高级通信，将客户机群和服务器群有机地“粘合”起来。

其工作流程是：在客户机里的应用程序需要驻留网络上某个服务器的数据或服务时，搜索此数据的 Client/Server 应用程序需访问中间件系统，该系统将查找数据源或服务，并在发送应用程序请求后重新打包响应，将其传回应用程序。

§ 3.2 系统结构设计：三层结构的图像检索模型

根据三层 Client/Server 结构的功能划分，我们很容易就能得到对应的三层结构的图像检索模型（图 3-2）。其中用户模块是用户交互和系统功能表示的接口，完成表示层的功能；功能层则由图像分析模块、知识库模块和交互查询模块构成，完成系统功能的应用逻辑；而数据库模块则提供数据服务。

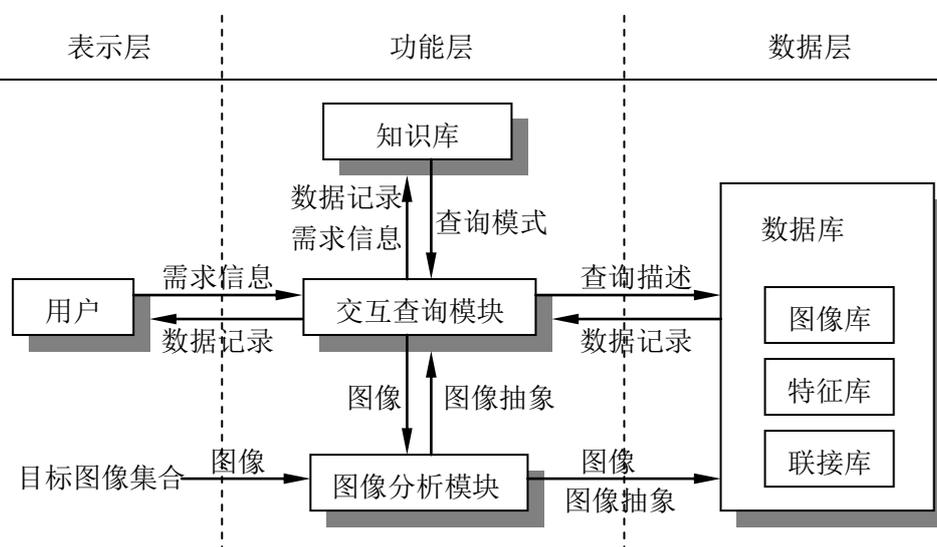


图 3-2 三层 Client/Server 结构的图像检索模型

但这种划分不是绝对的，在具体的实现中由于诸多因素可能需要作一定的妥协。比如为了能够减少网络传输量，我们可能需要在表示层实现部分的图像分析功能，这样当用户进行基于内容的图例查询时，我们可以直接在表示层得到图例的内容特征，可以只将表示这个特征的数据结构提交至功能层进行检索，而不是传送整幅图像。

有两方面的理由使得我们不能把知识库模块放到数据层实现，一方面，这里的“知识库”在实现时通常不以数据库的形式表达，而是一个包含特征的使用信息和图像相似比较的评价模型的集合，可以以不同的形式（程序、文件或规则）贯穿到功能层的应用逻辑中；另一方面，功能层的功能实现离不开知识库模块的有效支持，需要与其它模块紧密结合，以获得功能实现的效率。当然，在功能层中，知识库模块也可以作为规则库通过数据库来实现，但这时的数据库与我们前面论及的数据层的数据库服务无论从作用还是从内容上都是不同的，从这个意义上来说，知识库中的规则仍然是应用逻辑的规则，仍处于功能层。

图像检索模型中，所有的图像在插入数据库时都要通过图像分析模块，这时我们可以完成图像特征的预定义工作，包括各种基于内容的特征。在医学影像数据库系统中，图像分析模块处于一个十分重要的地位，它使得医学影像数据库具有了作为一种图像数据库应具有的特征和功能。图像分析和处理的性能将直接影响图像数据库的性能，它必须要有良好的结

构、较高的效率、较低的时空资源要求和相对完善的功能。本文将在下一章中对医学影像数据库系统中的图像处理问题进行讨论并为图像分析模块给出一个基于构件的实现。

§ 3.3 数据库设计

在医学影像数据库系统中，数据库处于核心地位，数据库的设计也成为设计工作的主要内容之一。从前面的论述中我们可以得知图像数据库应由图像库、特征库和联接库构成，我们在这里也按照这种划分给出相应的设计。

我们在第二章中给出图像检索模型时，为了使模型由更广泛的适应性，我们尽量避免与领域有关的分析与设计，这样的设计是与领域无关的。但对医学影像数据库，实际上是这一模型在医学领域的应用，所以在实际的设计工作中，应该充分考虑领域的需求。

这里我们考虑到三维医学可视化、法医人类学颅骨面貌复原等系统的实际需要，对数据库作相应的扩展，使之能够适应相应领域的需要。我们把这一类数据称之为领域库。领域库是完全构架在三个基本库的基础上的，我们同时给出一些与领域有关的库的设计，首先是出于应用的实际需要，同时据此也可以证实我们这种基于三种基本库的设计的可扩充性能，我们将看到，在三个基本库基础上针对应用作领域扩充是很容易的。

据此也可看到我们在数据库设计上的总体思路和路线（如图 3-3），从需求分析开始，我们将图像数据库分析和领域分析分别地但是相联系进行，前者的分析结果产生三个基本库，后者分析是在前者基础上进行的，其结果是构架于基本库之上的领域库。之后再行合并（消除冲突）和优化（消除冗余），给出的整体设计。

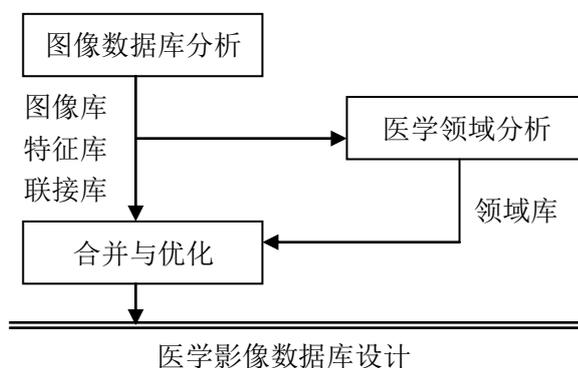


图 3-3 医学影像数据库的设计过程

我们用 ER 模型（Entity-Relationship Modeling）给出各类库的设计。我们的目标是在一种比较完备的关系数据库上实现图像数据库，所以设计完全是在关系规范理论的指导下进行的，并且考虑到存取效率和可维护性对某些关系进行了进一步的垂直和水平的划分。所有的关系将达到 4NF。

除了某些与医学有关的设计外，我们的设计可以适用于一般的图像数据库。

经过精心设计的医学影像数据库虽然在结构上非常清晰，而且也不复杂，但其存储要求可能十分巨大，在具体的实现中可以根据图像的实际存放位置、图像的存储形态、图像的用途等对数据库进行切分，并存放在不同的物理位置，进行局部和全局的管理和控制，以实现分布式的数据库（文献[7]）。

3.3.1 图像库

图像库是医学影像数据库的核心，是所有图像的集合和组织，并在数据库范围内被赋

予一个唯一的标识。图像在数据库中的存在形态可以有：

(1) 完整的有格式图像数据；

(2) 图像文件在网络内的完整路径名。系统可以通过这个路径名取得所指定的图像文件。当系统运行于 Internet 环境时，这个路径名将是图像文件的 URL。

第二种方式实际上只存储了对图像的引用，可以大大降低数据库的存储量。但同时也会引起数据库中关于图像的描述和特征信息与图像实际内容的不一致性，所以系统中应为这种存储设置定时检查机构，保证在一定的时间间隔内最大程度的一致性。

如果图像比较大，还应为图像建立统一格式与尺寸的压缩图标，以提高检索速度、降低网络消耗。

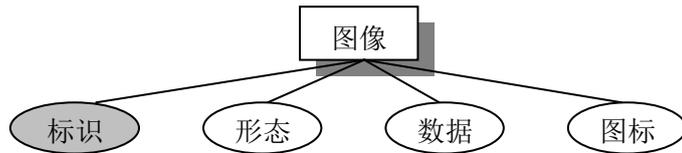


图 3-4 图像库 (灰色属性为码)

3.3.2 特征库

医学影像的特征分为三种，分别为图像结构特征、基于内容的特征和基于描述的特征。它们与图像的联系都是一一对应的 (如图 3-5)。这种一一对应的联系在关系数据库中可以很容易地用码之间的互为外码来实现。

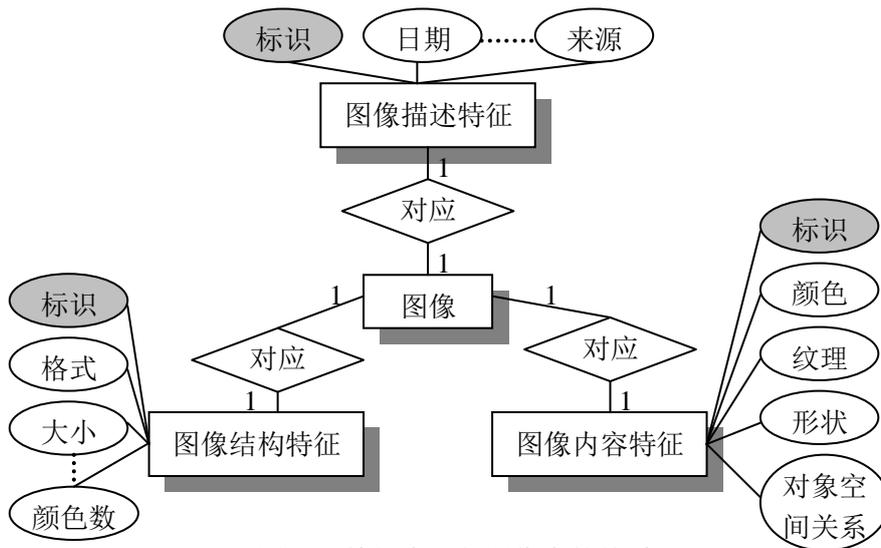


图 3-5 特征库及与图像库的关系

图像结构特征是指从结构化的图像数据中直接得到的一些信息，如图像的格式、尺寸、颜色数等，这些数据一般可以从图像数据的文件头中直接读到，将它们分别抽象出来是为了提高检索效率，没有必要为获得这些信息读取整个图像。基于内容的特征是指图像的颜色、纹理、形状和对象空间关系特征，这些特征经图像分析模块抽取后用结构化的二进制形式存入数据库。基于描述的特征反映了图像的最基本的一些描述性的信息，如图像的产生日期、产生方式（来源）、主要作用等。

3.3.3 联接库

联接库反映了图像与图像之间的联系，具有一定的领域特点。对医学影像数据库，我们通过对图像分组来抽象这种联系。图像总是属于某个图像组的，图像组可以通过与一些和

领域有关的信息反映图像与图像之间的联系，如对同一病人某次头部 CT 的所有影像的空间关系等。一般情况下，领域库必须通过图像组与图像取得联系，即便是一张单独的图片，也必须为他专门建立一个图片组。这样处理可以获得在结构、关系及处理上的一致性和简单性。

图像组至少应包含一幅图像，一幅图像可以因为不同的目的属于多个图像组，这种多对多的关系（图 3-6）可以为领域中图像的重复利用创造条件。

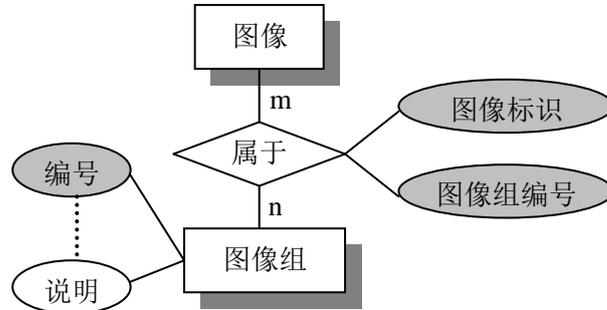


图 3-6 图像组是表达图像联接的一种方法

图片组内图像之间的具体联系取决于领域的定义。所以我们把图片组内的具体联接交由领域库定义。

当然，根据具体应用的不同，我们也可以定义其它形式的联接，比如图像之间的树形层次关系等。

3.3.4 领域库

目前我们的医学影像数据库系统涉及医院影像管理、医学三维可视化系统、计算机辅助的法医人类学颅骨面貌复员等应用，不同的领域对影像有着不同的观点。下面我们就这些应用领域中部分数据需求给出现相应的领域库。

(1) 脑外科 CT 切片库。

脑外科的 CT 切片是脑外科诊断与手术的依据。切片之间具有空间位置关系、比例关系和角度关系。切片之间的关系存在于属于同一个图像组的切片之间（图 3-7）。为保证切片组中含有的切片与图像组中的定义一致，可以在组内切片、图像与图像组的“属于”联系根据相同的码通过自然连接建立逻辑视图。用户对组内切片的定义和更新必须通过视图进行。

组内切片的这种关系是进行三维重构的基础，在我们前面所述的三个应用领域中都有

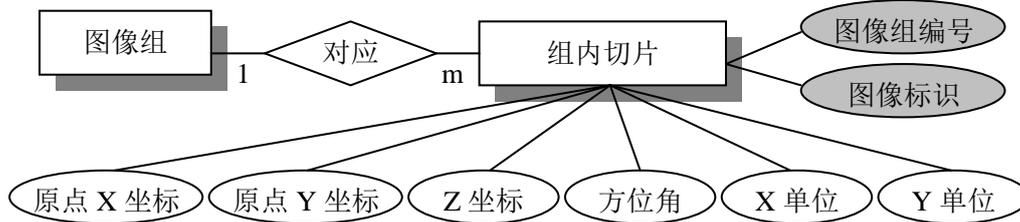


图 3-7 CT 切片组组内关系

重要的作用。

(2) 脑外科病例库。

根据病人的信息、影像产生或应用的科室、教学资料的结构等信息检索图像是一种常用的、有效的、为用户所熟悉和欢迎的检索方式。从另外一个角度看，通过应用领域的结构组织和管理图像也是用户所希望的一种方式，其本质是一种信息系统管理。关键是需要和图像数据建立联系，正如前面提到的，这类信息都要通过图像组与图片建立联系。我们以脑外科病例管理为例介绍这类联系方法（图 3-8）。

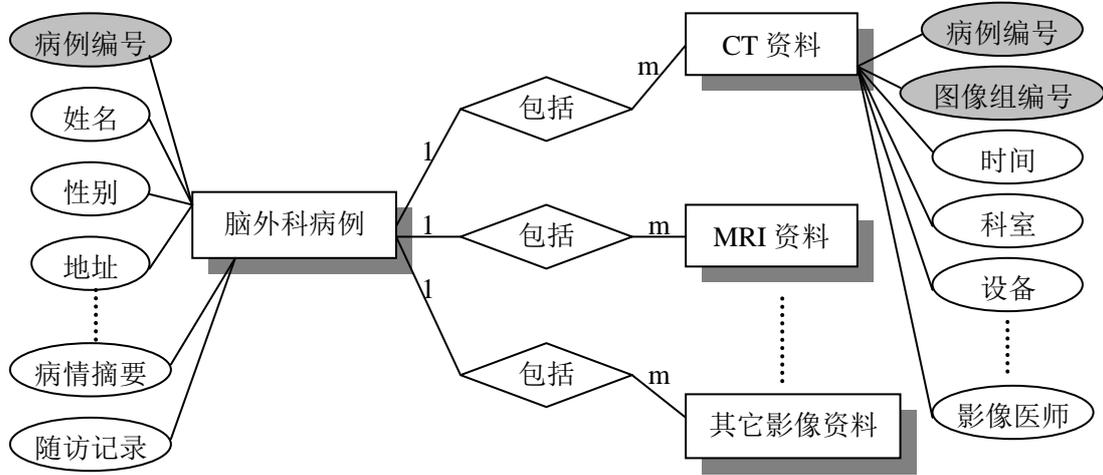


图 3-8 脑外科病例

(3) 三维体库

在医学影像的应用中，经常需要根据一组影像进行三维体的重构，重构的结果是一组三角形构成的三维体。医学影像数据不仅能为重构提供足够的切片信息，也可以为重构的结果提供存储。通过与三维可视化系统的接口，用户可以通过检索病人或图像而得到相应的重构三维体。体数据与图像数据的关系可以简单地用图 3-9 表示。

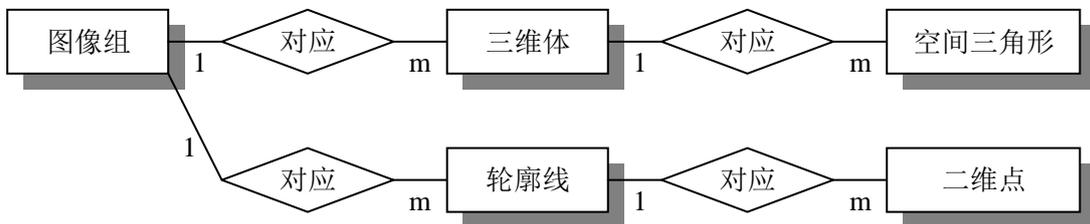


图 3-8 三维体与轮廓线也可存入医学影像数据库

(4) 轮廓线库

在计算机辅助的法医学颅骨面貌复原中，在三维重构之前，需要对二维切片数据进行表皮和颅骨的轮廓线的自动和半自动的提取，这些轮廓线将作为测量统计和三维复原的依据。所以有必要在影像数据中对这些轮廓线进行存储和管理。轮廓线库数据与图像数据的关系也在图 3-9 中简要给出。

从上面这些应用中可以看出，对医学影像进行分组是使图像数据有效地与领域相结合的关键所在，也使得图像数据库的结构简单明了，适应于在相同领域和不同领域中作进一步的灵活的扩展。

§ 3.4 远程图像存取中的问题与对策

3.4.1 问题

三层结构的医学影像数据库系统的目标是通过网络提供医学影像的检索和管理服务。基于远程网络的医学影像数据库系统能够为远程诊疗与教学提供重要的影像服务,然而远程存取时大多数情况下远程网络的有限带宽往往成为系统性能的瓶颈,尤其是图像数据的存取使得矛盾尤为突出。解决此问题的最根本的出路是加强基础网络建设以提供足够的物理网络带宽和通讯速度,但在此以前,我们必须从软件上予以充分的考虑,尽可能在现有网络基础上有效地实现医学影像数据库的远程服务。

3.4.2 多分辨率的检索结果

用户在检索数据库之前往往对要检索的图像只有一个模糊的、大略的概念,所以没有必要将所有的检索出来的图像完全显示给用户,用户往往只希望首先对检索结果集进行大致的浏览,发现真正需要的图像后才需要这些图像的精确表达。

这就需要对图像进行多分辨率的采样与编码。一般的数据库系统无法完成动态的、高效的图像处理,由应用系统预处理后将多分辨率的结果存入数据库又会大大增加数据库的存储负担。所以一个常用的折衷方案是,图像数据在经由功能层插入数据库时,同时由图像分析模块对图像进行缩小处理生成一个统一尺寸和格式的图像图标(Icon)(比如压缩成64×64JPEG格式的图标),同时插入数据库作为图像的一种微缩描述。在返回检索结果时,检索结果中的图像只返回对应的图标,用户可以通过图标获得对图像内容的初步了解,需要进一步的详细了解时,在由表示层提出请求取回相应的图像并显示给用户。

3.4.3 压缩存储

对图像进行压缩处理后可以极大地提高数据的传输效率。压缩的算法由多种,应根据图像的作用选择压缩算法和压缩参数,对一般的图像只需在保证视觉效果的前提下尽可能地压缩,但对用于医疗诊断或三维重构的医学影像则应以保证图像可用信息为前提,不能一味追求压缩比,甚至不允许带来稍许的信息损失。

下一章中我们将就医学影像的压缩问题进行更详细的讨论。

3.4.4 切分传送

我们这里所说的切分传输技术区别于网络的分帧传输技术,后者主要用均衡网络占用和差错控制,而前者主要针对数据库事务处理的粒度。

图像数据在传输过程中可能由于网络的冲突或故障引起传输的中断,虽然大多数的数据库系统可以对一个BLOB字段的操作作为一个单独的事务处理,也就是说事务的粒度可以达到单个记录的单个BLOB字段,但一般不能再小了。这使得当传输过程中发生中断时整个事务滚回,已经传输的部分图像将被视作无效数据而清除。

在更新数据库时,我们可以将图像数据在客户端进行切分,分批传送到数据库端后在数据库端合并,这样可以进一步细化数据库的事务处理力度,当传送发生故障时,事务的滚回只会引起当前片断的失效和重发,而不是整个图像的重发。

下面这段程序即是利用存储变量实现的图像切分更新(程序中略去了事务控制):

```
EXEC SQL BEGIN DECLARE SECTION;  
DECL_BINARY( 1024 ) buffer;
```

```
EXEC SQL END DECLARE SECTION;
EXEC SQL CREATE VARIABLE hold_blob LONG BINARY;
EXEC SQL SET hold_blob = "";
for(;;) {
    /* 将部分图像读入缓冲区 */
    size = fread( &(buffer.array), 1, 5000, fp );
    if( size <= 0 ) break;
    buffer.len = size;
    /* 在服务端逐步合并分段数据*/
    EXEC SQL SET hold_blob = hold_blob || :buffer;
}
EXEC SQL INSERT INTO some_table VALUES ( 1, hold_blob );
EXEC SQL DROP VARIABLE hold_blob;
```

切分传送不仅可以用于数据库的更新，也可以用于数据库的检索。在数据库系统中通过数据库服务将图像切分后，分批提取，在客户端获得足够的信息后将图像逐步显示给用户，用户可以尽早发现传送的图像是否是需要的图像并可以随时决定是否继续传输过程。

3.4.5 检索结果规模控制

控制检索结果集的规模可以有效地降低无谓的网络传输所带来的资源浪费。由于用户在检索图像数据时目标的不确定性，或者由于误操作，可能导致结果集的规模相当庞大，而且其中绝大部分并不是用户所需要的，这时就需要对检索结果集的规模进行控制，分批送回检索结果，交由用户确认后，再传送下一批结果。

通过控制结果集的规模是一种降低网络带宽浪费的有效手段，不仅是用于大规模的图像检索，也常见于各类基于文本的搜索引擎，在网络带宽十分宝贵的 Internet 上得到了广泛的应用。

在针对医学影像数据库的具体实现上，可以由功能层通过交互查询模块从表示层提交的检索需求中剥离检索条件，构造检索表达式向数据层请求符合条件的对象数目，若数目大于一定的程度，则重构用户检索需求，与数据层配合完成结果集的分割。此时应由功能层保留检索状态，以满足用户的对下一批结果的请求。这样做的缺点是为了完成规模控制需要数据层提供两次数据库检索服务。

对比较常见的结果类型，应在数据层建立相应的数据视图，通过数据库存储过程提供检索服务，并完成结果集的规模控制。这样可以进一步减少功能层和数据层之间的信息传输，并将上述的两次数据库检索降为一次检索，从而大大提高检索效率。但这种方法的灵活性和适应性也同时受到了限制。

3.4.6 在表示层实现部分图像分析

在前面所述的三层结构的图像检索模型中，医学影像数据库中所有的图像分析工作是由功能层的图像分析模块完成的，这使得功能层的图像分析和处理具有较高的独立性。但这种独立性有时也会带来额外的网络负担。

如果用户需要进行外部图例查询或草图查询，必须将图例发往功能层，由图像分析模块抽取其特征后再由交互查询模块构造有效的查询。如果能在表示层完成特征的抽取，那么就没有必要将示例图像送往功能层，而只传送表述特征的数据结构。

对此，我们对图像分析模块采用基于构件的嵌入式设计，一方面从另一个角度保证了图像分析模块的独立性，另一方面也使得图像分析模块获得了较高的灵活性，不再囿于功能

层,使一部分图像分析机制能够活动于表示层,从而降低了图像数据带给网络传输的压力。

3.4.7 缓冲技术

在功能层尤其是客户端表示层设立缓冲区,存放用户最近访问过的图像数据,可以使这些图像在很短的时间内重新被访问时无需从远程数据库中提取。一方面可以提高系统的相应性能,另一方面也可以在一定程度上降低网络传输量。

但要注意的是,对不同作用的医学影像应设置不同的失效时间。在本地缓冲中的用于临床诊疗的医学图像应尽快失效淘汰获得更新。而用于教学目的的图像则相对稳定,在本地缓冲中可以有较强的有效时间。

而功能层的缓冲由于访问频繁,可以设立较大的缓冲以满足众多用户进程的同时访问,同时为了保证缓冲区中重要图像的有效性,应在对这类图像进行数据库更新时,同时跟新缓冲中的图像,即采用所谓的“写通”技术。

§ 3.5 开发与运行环境

由于我们对医学影像数据系统采用灵活的三层 Client/Server 的结构设计,各层任务明确,功能独立,使得我们在不同的层次上可以从容地在多种开发与运行的环境中做出选择。

3.5.1 开发环境

硬件环境: Pentium 微机, TCP/IP 网络;

操作系统: Windows 95, Windows 98, Windows NT;

Web 前端: Allaire Homesite 2.5, Microsoft Frontpage 98, Borland C++ Builder 3;

GUI 前端: Borland C++ Builder 3, Microsoft Visual C++ 5.0;

中间件: Borland C++ Builder 3, Microsoft Visual C++ 5.0, Allaire Cold Fusion 3.0;

Web 服务: Microsoft IIS, O'Reilly Website 1.1;

数据库服务: Sybase SQL Anywhere 5.0。

3.5.2 目标系统运行环境

Web 服务操作系统: Microsoft Windows NT, Sun Solaris;

Web 服务: 几乎所有上述两种操作系统支持的 Web 服务

GUI 前端: Windows 95, Windows 98, Windows NT;

Web 前端: 几乎所有的浏览器(需同时支持 JavaApplet、JavaScript、ActiveX);

数据库服务: 几乎所有的完备的关系数据库系统。如 Sybase 10&11、Oracle 7&8、SQL Server 6&7、Borland InterBase 5.0 以及其它所有支持 ODBC 的比较完备的数据库产品,如 Access 97、Paradox 7 等。

建议采用的环境是:

Web 服务: Windows NT+Microsoft IIS;

GUI 前端: Windows 98;

Web 前端: Microsoft IE 4;

数据库服务: Sybase SQL Server 11;

§ 3.6 小结

本章中我们首先考察了三层 Client/Server 的数据库系统的基本硬件结构、各层的功能分配和这种结构的优势,认为这种结构非常适合于医学影像数据系统,随后,我们将医学影像管理需求和图像检索模型在三层 Client/Server 结构上进行了贯彻与落实,很容易地在第二章的基础上实现了影像数据库应用系统的总体结构设计。

数据库设计一直是数据库应用系统的核心,由于医学影像数据库系统是与医学领域紧密结合的,所以我们在数据库设计是对第二章中的数据库模型进行了扩展,在原有的图像库、特征库、联接库的基础上,适当缩小了特征库和联接库的范畴,增加了领域库以描述与医学领域相关的信息。这种收放后的数据库结构不仅适合医学领域信息的进一步扩充,也使得这种模式可以在实现其它领域的图像数据库时有很高的参考价值。要注意的时,这里的“库”并不是一个独立的数据库,而是医学影像数据库的一种根据功能和作用的不同所作的一种逻辑上的划分。

医学影像数据库必须要面对的一个问题是远程网络的有限带宽将在图像存取时显得捉襟见肘。为此我们从软件的设计与实现方面提出了一些相应的对策。

在本章的最后我们还给出了医学影像数据库系统的开发与运行环境。这是我们在本文中唯一涉及开发与运行平台的地方,因为整个系统分析与设计伊始,我们就希望我们的技术与成果能够具有最大程度的平台无关性和可移植性,尽管在最终进行原型系统的实现时我们不得不在这方面做出一定的妥协,但从目标系统的运行环境中可以看出,我们的努力还是有所收获的。

第四章 医学影像数据库中的图像处理

虽然目前有很多著名的商用图像处理软件如 Adobe® PhotoShop®、甚至一些免费而且功能强大的图像处理软件如 GNU 通用公共许可下的 The GIMP 等，都可以用来对医学影像进行处理。这里我们的目的不是在我们的医学影像数据库系统中再实现一个类似的系统，这是完全没有必要的，而且也是不切实际的。

之所以要在医学影像数据库系统中有机地集成一些图像处理功能，主要是出于这样的一些考虑：

(1) 医学影像数据库系统出于特征抽象的需要，必须在系统内部有效地提供无缝的图像分析与处理支持。

(2) 医学影像数据库系统的用户都不是图像处理专家，往往不能有效地使用复杂的通用图像处理系统，他们只需要一些常用的适合医学影像的处理手段和技术。

(3) 通用图像处理系统可以在将医学影像看作一般的图像的前提下进行它所支持的所有处理，但并不是所有的医学影像需要的处理它都支持。医学影像往往需要进行一些与应用和领域有关的特殊分析与处理，如根据医学影像的附加信息对器官和病变进行自动或半自动的精确测量等。

(4) 图像数据库系统必须达到一定程度的可用性。用户没有必要为了完成简单的图像处理工作在数据库系统和图像处理系统之间频繁地进行来回的切换，对医学影像进行格式转换、压缩和处理分析，所有这些工作通常应有系统自动的或者在用户的指导下半自动地进行。

(5) 在网络环境下，将图像处理功能集成到不同层次的数据库系统服务上，可以根据需要进行“就地”处理，而没有必要都传送到客户应用端处理，可以有效地降低传输开销。当医学影像数据库系统运行于远程网络时，传输开销将是影响系统性能的最重要的因素之一。

§ 4.1 医学影像的格式

正如绪论中指出的那样，医学影像数据库必须面对的一个重要问题是影像的格式问题。进行医学图像数据的处理，首先需解决数据来源。一些最新的高档医学图像设备本身已经与计算机结合在一起，而且往往必须由内含的微处理器进行信息处理。X-CT 和 MRI 机就是明证。如果我们需要把病人的检查结果存入电子病历数据库，只需从设备的接口上传出图像数据。当然，各种设备、图像数据的格式可能不同。没有统一的标准，这给医学图像数据的传输和使用带来了严重的问题。所以，在标准问题上，10 多年来有关专家做了巨大的努力。1993 年推出的医学数字图像和通信标准 DICOM 3.0(Digital Imaging and Communication in Medicine)是医学图像和相关数字信息在计算机间通信的一个工业标准。医学图像设备一旦采用了此标准，其图像的数据格式和数据接口就可在目前通用的国际联网系统上存储和传输。这是一个通用的标准，能满足高速传输图像、文字、表格、数据、动态图像以及声音的需要，而对某种实际应用，也许只需要其兼容子集，可在具体实现中予以简化。但为了便于与其他计算机软件相兼容配合，仍需要将 DICOM 转换为通用的标准的数字图像格式。

遗憾的是，我国一般医院目前尚无力以巨大投资更新各种昂贵的医学图像设备。即使更新设备，也不大注意设备与计算机接口的需要，而往往造成无法直接从医学图像设备向计算机传递有效的标准图像数据。这给医学图像信息的收集和存储带来了困难，其信息无法进一

步有效利用。

在目前大量使用的医学图像设备中,有些还处于非数字化状态,甚至有些还不是可由电子屏幕显示的视频信息,而仅仅是以肉眼观察的光学仪器。这就必须通过一定的手段来采集和转换,使之成为计算机能接受和处理的数字图像信息。借助摄像头和采集卡是一种常用的图像采集手段。根据不同图像对分辨率、黑白或彩色、静态或动态等的不同要求,选择适合的摄像头和采集卡便可实现图像采集。处理已有的 X 射线照片(或 X-CT 扫描断层胶片)可用高分辨率多灰度的单色扫描仪来输入。解剖样品卡片也可用彩色扫描仪来输入。也有一些医学图像设备,例如,某些型号的电子内窥镜或彩色多普勒超声波扫描仪等装置,虽不能直接提供数字图像信息输出,但可提供视频信号(或录像带),也可通过采集卡输入到计算机中。

所以在我们的医学影像数据库系统中,对多种专用医学影像格式和通用数字图像格式提供了支持,不仅支持这些格式的编码与解码,同时也支持这些格式间的相互转换。

对专用医学影像格式,某些较新式的医疗设备提供了专门的软件向一些流行的通用格式进行转换,而国内大多数的医疗机构尤其是我们的协作单位的医学影像设备由于生产商出于商业上的原因都采用了非常封闭的甚至经过加密的图像格式,为了有效地利用这些的像资源,拓宽医学影像数据库系统的适用范围,对这些图像进行分析和解码虽然很困难,但也是必要的。在这方面西北大学可视化技术研究所作了大量的艰苦工作,在医学影像数据库研究中涉及到的专用医学影像格式方面的问题的解决就是建立在这些工作基础之上,

通过专门的接口程序将这些专用格式的医学影像转换为标准格式后,医学影像数据库的主要任务便是对多种标准图像格式在数据库服务的功能层和表示层上提供支持。我们为医学影像数据库设计构造了基于构件的嵌入式图像处理系统,能够对目前流行的大多数的数字图像格式提供编码、解码、压缩和两两之间的转换。这样用户在数据库系统的应用中就可以得到图像格式上最大的灵活性,用户只需向系统提供用户认为有价值的图像,而不必受到格式上的限制,而且相反,用户可以对同一幅图像要求数据库系统提供不同格式的表达,以用于不同的目的和场合。

目前我们的原型系统已经可以处理的图像格式有 BMP, PCX, JPEG, GIF, TIFF, WMF, EMF, ICO, CUR 等,基本上满足了应用的需要。

§ 4.2 医学影像的压缩

现代化的医院管理和远程诊断的基础就是要建立电子病历,将有关患者病情的声、图文并茂的多媒体信息记录在病人档案数据库中。首先遇到的问题就是医学多媒体信息数据的存储问题。由于图像信息量较大,为了便于存储和传输,通常采用压缩的手段。有的采集卡在采集的同时就能进行压缩,尤其对高分辨率的彩色图像更有必要压缩。

由于在存储和传输上的需要,对医学图像进行压缩是很有必要的。由于医学图像不同于一般的用于观赏或示意的图像,它主要有两方面的应用,一是治疗诊断目的,必须保证足够的信息用于医生的诊断、计算机的后续处理、甚至进行手术计划与导航,信息的稍许丧失可能会导致严重的后果;二是用于教学研究目的,只要能进行示意或显示,表达一定的主题信息即可,在这个意义上它已一般的图像类似。所以对医学影像的压缩应由用户针对其目的,分别采用无损或有损的压缩方法。无损压缩的算法删除的仅仅是冗余的信息,因此可以在解压缩的时候精确恢复图像,有损压缩的算法把不相干的信息也删除了,因此只能对原图像进行近似的重构,而不是精确的恢复。我们容易得知,有损压缩的算法可以达到较高的压缩比。对于多数图像来说,为了得到更高的压缩比,保真度的轻微损失是可以令人接受的折衷方法。而有些图像,尤其是治疗诊断用的图像,更多的应用场合下不允许进行任何的修改,在这些

情况下,只能对他们进行无损压缩。

4.2.1 无损压缩技术

无损数据压缩算法可以分为两大类:基于字典的技术和基于统计的方法。基于字典的技术生成的文件包含的是定长码(通常是12至16位),每个码代表原文件中数据的一个特定序列。基于统计的方法通过用较短代码代表频繁出现的字符,用较长代码代表不常出现的字符,从而实现数据的压缩。

基于字典的技术主要有行程编码(RLE)和有专利保护的LZW算法。统计编码方法中较常用的是哈夫曼编码(Huffman Coding)(文献[8])。

西北大学可视化技术研究所的谷东伟曾尝试用二叉树的方法进行无损的图像压缩,在医学可视化分析平台上得到了较好的应用。

4.2.2 图像有损编码

图像有损编码可以采用标量量化或变换编码的方法。所谓的标量量化就是将图像量化成较少的几个灰度级以减少数据量。而图像变换的方法是对图像进行变换,删去那些接近于零的系数,对那些较小系数进行粗量化,由此将传输和存储的数据集中到那些包含图像主要信息的系数上,在以后重建图像的时候,那些重要的内容很少会有损失。常用的基于特征向量的图像变换编码技术有SVD变换、K-L变换、DCT变换、方波变换、小波变换等(文献[2][8][10])。

有些压缩方法则是上述一些方法的综合。如由国际标准化组织(ISO)和CCITT联合发起的联合图像专家组(JPEG),建立了静态图像压缩的公开算法,首先将RGB分量转化为亮度和色差分量,同时丢失一半的色彩信息(空间分辨率),然后用离散余弦变换(DCT)来进行块变换编码,舍弃高频的系数,并对余下的系数进行量化以进一步减少数据量。最后使用RLE和Huffman编码来完成压缩任务。JPEG的解压缩过程就是JPEG压缩过程的逆过程,这时的算法具有对称性。在视觉效果不受到严重损失的前提下,算法可以达到15或20的压缩比,如果在图像质量上稍微牺牲一点的话,可以达到40:1或更高的压缩比。

4.2.3 在医学影像上的压缩实现

由于我们将多格式的医学影像首先转换成标准的图像格式,这些标准的图像格式中有相当一部分支持各种形式的压缩,可以满足医学影像的需要,所以在本文涉及的工作中,压缩工作与格式转换工作是同时进行的。由于系统兼容性和Internet传输和显示的需要,虽然我们实现了多种图像格式和压缩方式,但我们推荐用户采用BMP或JPEG格式,这些格式的图像可以被大多数的WWW浏览器所支持,且前者可以支持无损的RLE压缩,后者可以根据需要调整压缩参数和算法进行高压缩比的压缩。

§ 4.3 医学影像的处理

医学影像本身以普通图像的形式存在,原则上说处理普通图像的方法都可以适用于医学影像,但由于医学影像具有一定的科学性和完整性,根据处理后图像的用处不同可以有不同的处理原则。一般说来,用于医疗诊断的图像的处理原则应该是不能在医学影像所反映的信息基础上人为地“制造”新的信息,而是使图像更好地反映其原有信息。

实际工作中我们选择用户提出的常用的处理进行了实现。

4.3.1 颜色调整

由于医学影像在采集、存放过程中的问题，有可能在颜色上产生偏差。如以前医疗机构或条件不完备的医院，其医学影像大都以胶片、照片或生物切片样本的形式存放，由于时间、环境或气候原因使得图片的色彩质量受到了影响。为了建立新的电子档案或进行远程诊疗，可能将这些图片通过扫描仪数字化。此时得到的医学图像就需要进行颜色调整。

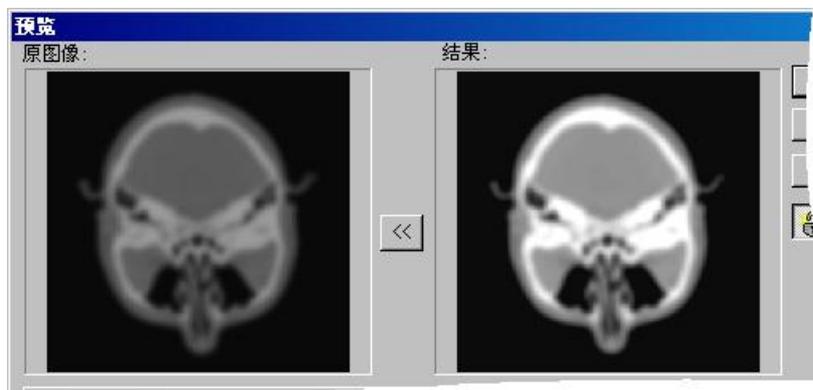


图 4-1 对比度和亮度调整前后

颜色调整包括亮度、饱和度、色调、对比度的调节，颜色模式的转换，颜色的色标转换等。图 4-1 便是对一幅对比度较弱的图片进行对比度和亮度调整前后的质量对比。

4.3.2 图像锐化

医生在进行影像分析与诊断时，经常要测量某些器官或病变的宽、高、周长、面积等。如果在做这些度量之前先加强图像中器官或病变的边缘和轮廓使得度量更加方便可靠，也可以使得自动测量成为可能，这就需要进行图像锐化。

图 4-2 给出了使用一个一般的 3×3 卷积窗口进行边缘检测的例子。

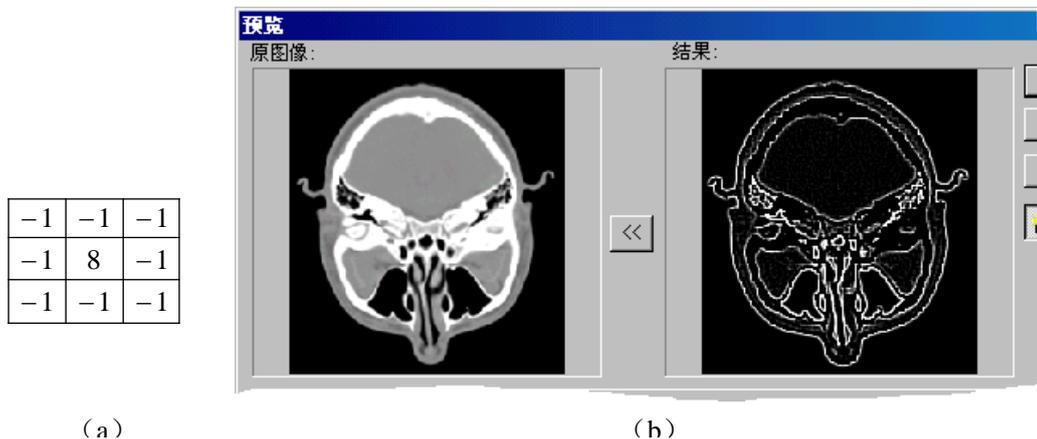


图 4-2 边缘检测的 (a) 卷积窗口；(b) 效果

4.3.3 图像增强

图像增强的目的是使得图像看上去跟清晰。我们在工作中采用像素强调卷积寻找图像的细节并强调它们，也即用像素强调窗口作图像卷积通过高通滤波的方法来清晰图像。图像的增强一直是学者们研究的热点之一，有很多更优越更复杂的方法通过各种途径来增强图像的清晰程度，尤其是医学图像的增强在应用中有较广泛的需求。

西北大学可视化技术研究所的刘硕专门就医学图像的图像增强问题作了较广泛和深入的研究，并提出了一种采用模糊数学的方法增强图像的方法，收到了较好的应用效果（文献[31]）。

图 4-3 给出了采用像素突出 3×3 卷积窗口清晰图像的示例。

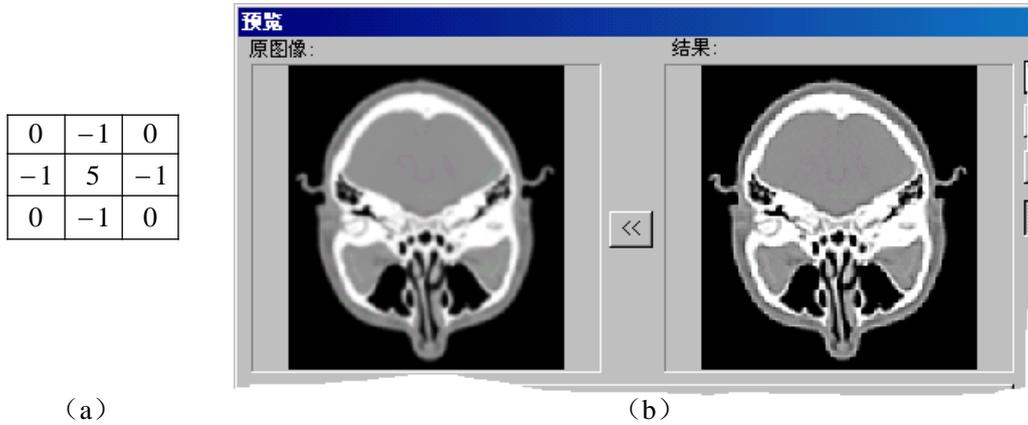


图 4-3 像素突出的 (a) 卷积窗口；(b) 效果

4.3.4 图像均衡与直方图调整

严格说来，图像均衡与直方图调整也属于颜色调整的范畴，但由于在应用上常独立运用，所以在此专门作一说明。

有时医学图像以胶片或照片的形式存放时间过长而颜色发黄，或者采用照相或扫描的方法在进行图像采集时采用了不恰当的光照，如在室内采用了日光型的胶卷，或曝光过度等，这时会出现色彩不平衡的问题，就需要进行颜色的均衡。对从数字医学影像设备如 CT 机、磁共振设备等采集的灰度图像，有时为了观察效果也需要进行灰度均衡。

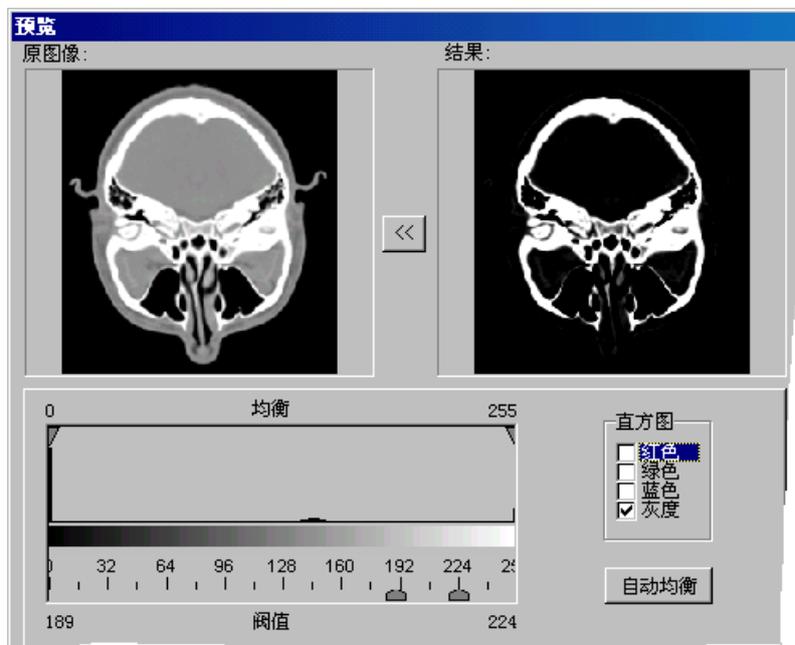


图 4-4 直方图调整

直方图调整是一种手工均衡的方法，通过在 RGB 颜色通道和灰度通道上调整均衡参数（均衡区间）来调整图像的效果。通过直方图调整的方法也可以在 RGB 颜色通道和灰度通

道上设置过滤阈值区间,使得呈现某种灰度或颜色区间特征的组织或器官在图像中突显出来,以满足某种医学上的需要。在国家“九五”重点科技攻关项目“计算机辅助的法医人类学颅骨面貌复原”中,就需要根据一般的CT切片突现颅骨形状,以便于确定颅骨形状和三维重构。图4-4就给出了基于这种需求的直方图灰度调整的一个示例。

4.3.5 缩放

在医学影像数据库系统中,支持图像的有效缩放是很有必要的。因为用户在浏览检索结果时通常首先需要在有限的屏幕空间中尽可能多地观察满足检索条件的图像,在发现目标图像后有需要仔细地观察某幅图像或图像的一部分。前者需要缩小图像,而后这需要放大图像。而且为了支持多分辨率的存储和表达,缩小处理也是必需的。

这里我们使用了“有效缩放”一词,主要有两方面的含义。首先是信息有效性,放大和缩小都应尽可能多地保持图像在原尺寸下所表达的信息。其次是用户有效性,应支持用户根据其需要进行任意比例的缩放和存储,由于使用非常频繁,缩放操作操作应简洁流畅。用户有效性通过精心设计的操作界面是比较容易达到的。这里我们主要讨论信息有效性。

放大图像最简单的方法是复制每个像素数次,但在视觉效果上比较粗糙,呈颗粒与锯齿状。解决这个问题的比较好的方法是插值的方法,可以采用相邻两点的线性插值,也可采用相邻3个点以上的非线性高次插值。前者可以获得较高的速度,而后者是以较高的运算代价换取较好的效果。为了能够进行实时的放大,而且保持较好的效果,我们在实际工作中采用了卷积技术,也即先用复制方法放大图像,然后用一个平均窗口对像素复制图像作卷积,平均窗口的大小等于放大系数,窗口的权都一样并令所有权的和为1。如对于放大4倍的图像,卷积窗口应为所有元素都是0.0625的4×4的矩阵。

缩小一幅图像最简单的方法是简单地每隔N个点抽取一个像素,也即“抽点”的方法。但这样会引起空间频率混叠(Aliasing),可能会使许多重要的信息丢失。我们知道,至少要用图像最高频率的两倍频率采样,才能真实地捕获图像的要素,并且不会引入混叠频率(采样定理,文献[2][8])。所以在缩小应用中,应先用低通滤波器过滤源图像,使得图像中不包含高于1/2采样频率的空间频率分量,然后每隔N个像素取一个像素点,这样处理的图像就不会有混叠效应出现。

应用中我们采用了像素平均化图像缩小技术。先用低通滤波窗口对原始图像作卷积,窗口的大小应该是图像缩小的系数(对缩小系数为9,窗口应为9×9),如果出于效率的考虑,可以采用平均滤波窗口,即平均窗口的大小等于缩小系数,窗口的权都一样并令所有权的和为1。滤波完成后则按照需要抽点,形成缩小的图像。

4.3.6 其他处理

从上面的叙述可以看出,除了颜色调整以外的其他处理我们主要采用卷积滤波技术。我们同时将3×3窗口的卷积滤波作为一种通用的手段设计成一种处理接口,以便于扩充处理功能。我们可以通过这个接口随时为用户(也可以由用户自己)设计出理想的滤波器(如图4-5)。

通过预定义的或自定义的滤波器参数,用户还可以完成诸如各类高通或低通滤波、去卷积、去噪声、特征增强等图像处理工作。



图 4-5 预定义的或自定义的滤波器

§ 4.4 基于构件的嵌入式处理系统设计

即便我们采用了灵活的处理接口以满足不同的医学影像处理需要,但仍然不可能完全估计到目前和将来用户在医学影像处理方面的需求。而且这里我们的主要任务是为医学影像数据库系统提供图像分析模块所需要的处理功能,和作为用户接口的一部分向用户提供数据库数据采集和检索时必要的图像处理接口,我们不可能也没有必要为此设计一个功能强大而且完善的医学影像图像处理器,这似乎应该是另一个专门课题的研究任务。

但为了能够适应这种不断变化和增长的处理需要设计一种具有高度可扩充性的处理系统结构和机制是必要的,以使得在新的处理需要或者新的处理技术出现时处理系统能够迅速得到扩充,而且不影响数据库系统的其他部分。

那么,这个图像处理系统应该具有这样一些性质:

(1) 嵌入特性。图像处理系统作为一个整体可以嵌入到医学影像数据库系统所有需要图像处理的地方。

(2) 格式无关性。它接受的输入和能够完成的输出是数据库可以采集、存储和显示的所有格式的图像。

(3) 接口可视性。在数据库系统需要进行交互式处理的人机界面上提供处理过程和方法的交互式控制,以及中间结果及最终结果的交互式显示。

(4) 处理方法易扩充性。可以很方便地扩充处理能力和处理能力,而不需要对处理系统的其他部分,以及处理系统以外的数据库系统的其他部分进行修改。

为了实现这样一个医学影像处理系统,我们采用构件技术,将整个处理系统封装成一个构件,而这个构件又由多个功能独立的多个子构件构成。从而分别完成上述的各项性能指标(如图 4-6)。

处理系统构件主要由以下几部分组成:

(1) 数据库存取部件,完成图像数据的数据库存取;

(2) 图像 I/O 部件,进行各种数据格式的解码、压缩、编码和格式间转换,对输入的外部格式图像自动判别其格式,形成并接受统一的内部格式图像数据流,以实现格式独立性。

(3) 图像处理工具箱,对内部格式的图像数据进行各种处理,处理方法的改进和扩充只通过这个构件进行,以使处理方法实现易扩充性。

(4) I/O 参数调整工具箱,含有多种可重组可视控件。负责与用户交流格式信息,用户可以通过这个控件选择图像格式,指定编码方式,选择压缩算法,调整压缩参数并了解当前 I/O 参数。与下述两种可视控件一起完成各种处理接口的可视性。

(5) 处理参数调整工具箱,含有多种可重组可视控件。用户可以通过这些控件了解处

理的各种参数和状态，选择处理方法，设置处理参数，控制处理的进程。

(6) 图像显示控件，含有多种可重组可视控件。一方面向用户提供图像显示服务，另一方面用户可以对显示进行控制、编辑和图像输入，并将用户的操作作用于内部数据流。

整个处理系统可作为独立构件运用于医学影像数据库系统中。需要指出的是，如果系统不需要用户进行可视交互，也即图 5-6 中的“用户可视界面”变为一般用户（程序）界面时，所有的可视控件可以将自己隐藏起来作为一般的构件提供服务，交互将在程序之间进行。

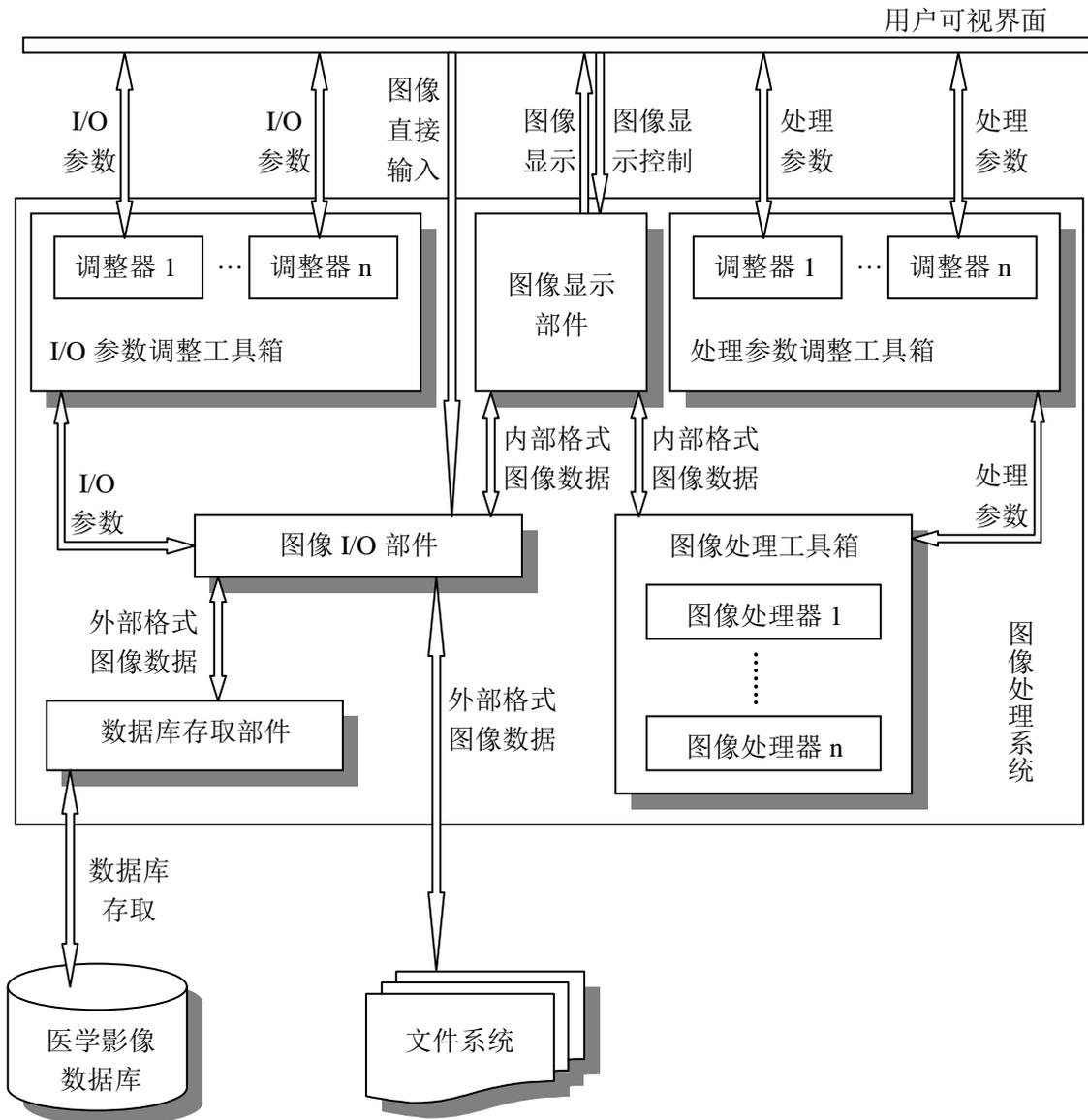


图 5-6 医学影像数据库系统的图像处理系统构件

由于整个处理系统构件时通过各个子构件聚类而成，各个子构件也可以嵌入到数据库系统中不同的位置各自独立活动。这样可以有利于图像处理机构的不同部分活动于数据库服务的不同层次上（甚至不同的物理位置上），从而获得较高的灵活性。

§ 4.5 小结

在本章中我们讨论了在医学影像数据库系统中有机地集成一些图像处理功能的必要性，

并对医学影像涉及到的格式转换、数据压缩和处理方面的部分问题进行了分析和讨论。最后还给出了一个用于图像数据库系统的基于构件的嵌入式图像处理系统模型和结构。

当然，本章论及的图像处理系统并不是排斥通用图像处理软件的使用。我们认为那些对医学影像来说不常用的，但在通用图像处理软件已经有完善实现的，应鼓励用户通过通用图像处理软件处理，毕竟影像数据库集成的图像处理功能不是影像数据库的主要目标，这里我们只能提供**必要的**而不是全部的图像处理功能。

第五章 基于内容的图像检索技术 (CBIR)

早期图像数据库的检索方法主要是基于描述的直接查询 (Direct Query By Description), 通过对每一幅图像建立关键词等描述信息作为图像索引, 采用传统数据库技术就可以满足图像检索要求。但是一些图像的文本信息难以描述, 不同人对同一幅图像有不同的理解, 这样文本描述信息就存在着二义性, 于是人们就提出了基于内容的图像检索方法。

基于内容的图像检索 (CBIR, Content Based Image Retrieval) 或基于图像内容的查询 (QBIC, Query By Image Content) 技术是指除了利用传统的数据库对与图像有关的文字信息进行存储管理外, 还要利用图像的颜色、形状、纹理、对象空间关系等特征对图像进行查询。这一新的研究方向涉及图像处理 (Image Processing)、图像识别 (Image Recognition) 和数据库技术 (Image Database), 把这三个领域的研究成果进行有效的结合是该技术取得成功的基础。

基于内容的图像检索技术由于能够根据图像的可视内容产生查询, 从而方便了用户, 因此在许多领域具有广泛的应用前景。除了用于医学图像处理与材料分析, 诸如艺术馆和博物馆馆藏资料的管理、地球资源遥感图像的检索、建筑和工程设计图纸的查询、时装设计、保安电子化以及犯罪记录调查等, 尤其是计算机网络技术和通信技术的发展, 未来在网络购物和交互式电视等方面, 这项技术大有用武之地。此外, 多媒体系统还将以惊人的速度深入到家庭生活中, 如应用到生活照片的管理中。因此, 它的开发和应用引起了人们广泛的兴趣。

本章中我们有必要对基于内容的图像检索技术在总体上作一综述, 更重要的是, 我们将提出一个解决 CBIR 问题的总体技术路线。下一章我们将把这一技术路线贯彻到颜色特征上并给出详细的论述。

§ 5.1 CBIR 的技术特点

基于内容的图像检索技术的主要特点有:

首先, 它是一种信息检索技术, 应能从大型分布式数据库中以用户可以接受的响应时间及尽量与领域无关的 (在需要时也可以是与领域相关的) 方式检索到想要的图像信息。它可以不去理解图像中的对象, 更关注的是信息的快速查询和发现。举个例子来说, 用户想买一辆红色汽车, 那么他选定了红色和汽车轮廓, 数据库系统就能返回所有红色汽车的图像, 从而帮助用户快速找到目标。

其次, 作为一种多媒体技术, 它具有很强的交互性, 即用户可以参与检索过程。

最后, 它引入了特征库和知识辅助的概念。用特征库保存描述图像内容的特征, 知识辅助有利于查询优化和快速匹配。

从图像检索的应用角度出发, 除了采用传统的结构化查询方法以外, 还可以采用可视的示例查询方式。不论哪种查询方法, 都需要针对图像的具体特征进行匹配检索。

§ 5.2 CBIR 的一般过程

基于特征的图像检索过程一般有以下几个步骤:

(1) 图像的预处理

主要工作有图像去噪音、图像尺寸的统一和图像的旋转等, 这些工作是图像特征抽取

的基础。去噪音时可将彩色图像按三原色分别做滤波处理,并将所得结果合并输出。另外,滤波处理还可降低背景信息的影响,滤波器一般具有均值滤波和中值滤波两种方式。

(2) 图像特征的抽取

图像特征抽取是指从包含大量信息的图像中分解出不同种类的特征。图像的特征大致分为两类,即人的视觉可以直接感受到的自然特征和人眼无法直接观察到的特征。前者被称为图像的视觉特征,如二值化信息、区域的边缘轮廓、纹理、亮度和清晰度等;后者被称为图像的统计特征,如图像的直方图和频谱等。

(3) 图像库的搜索

常用的图像搜索策略有自适应聚类搜索、模糊搜索、图像的文字信息与图像信息混合搜索等。

§ 5.3 一个解决 CBIR 问题的总体技术路线

目前,国内外有很多的学者就 CBIR 问题进行了许多有益的探索,并已经取得了一些成果。通过我们所掌握的技术资料,纵观这些 CBIR 方面的方法和技术,可以说是见仁见智,各有春秋,创造性地从各个方向通过各种技术手段来解决 CBIR 问题。但可以看到,这些研究往往是针对 CBIR 中的某一方面如颜色(文献[3][4][5])、纹理(文献[11][20])、形状(文献[6][19])或空间关系(文献[19])展开的,使得这些在视觉上相联系的特征在研究中被孤立出来。当然,在某些特定的问题环境下,可能某种特征是主要的,甚至是唯一被关心的,但在大多数的情况下,用户往往更希望通过一种简单的、一致的、多角度的方式进行 CBIR 的应用。由于孤立地进行基于某种特征的研究,虽然在不同的方向上取得了不同程度的进展,事实上能够均衡考虑各种特征、在时间和空间代价上可行、真正投入运行并得到用户广泛认可的系统并不多见,甚至是屈指可数的。

我们在开展医学影像数据库的研究的同时,并不囿于专门领域的狭隘观点,同时把寻求一种有效的具有普适性的 CBIR 检索机制和技术路线作为我们的研究目标,然后再把这一路线在充分考虑领域的特点和需求的前提下贯彻到医学影像数据库的研究和实践中。

简单地说,我们解决 CBIR 问题的总体思路是:用一种一致的、紧凑的、自然的、完整的方法来看待和表达图像的颜色、纹理和形状特征,有效地表达图像中对象的空间关系,将这些特征和关系通过预定义的方法存储于数据库中用于索引和检索图像。表达应该是简单有效的,有较少的时空代价,使得我们可以在现有的完备关系数据库基础上实现图像检索。

5.3.1 基于二值直方图的特征表达

特征表达的主要任务是表达某种特征在图像中的分布情况,也就是要指出是否含有某种特征的元素以及具有这种特征的元素在图像中所占的比例。在应用中主要目标是回答用户在相似意义上的“是否含有”、“含有多少”的逻辑。特征表达与图像的大小、角度和对象位置无关。

对颜色、纹理和形状这些视觉特征来说,一个图象点(Image Points)可以被映射或变换 T 为多维特征空间(Feature Space)的特征点(Feature Points),对特征空间进行量化 Q ,从而获得一维的特征分布描述。

使用这些特征元素,我们将应用两种方法来表达特征:直方图和二值集。直方图(Histogram)是特征元素的一种分布,用直方图的方法表达特征具有缩放、旋转和平移无关性(在变换不引起特征的损失的前提下)。而二值集(Binary Set)是特征元素直方图表示的一种逼近和压缩,是图像中较重要的(对应较多的图像点)特征的表达,使得我们能把注意力集

中到视觉特征上去,如主色调、大片的纹理、占统治地位的形状等,而且我们能以较小的空间代价描述特征分布。这种描述是近似的,但能够满足特征表达的需要。

三种不同的视觉特征的特征空间变换 T 是不同的,但量化过程 Q 和结果、以及通过特征空间中距离 D 进行相似性度量的过程和算法基本上是一致的,使得一旦获得了图像的特征分布表述,我们就可以以一致的方法表达特征并进行相似性度量。

下一章我们将以颜色特征为例,详细论述这一过程。由于量化过程 Q 和相似性度量基本一致,所以不再给出纹理和形状的特征空间量化和相似性度量。而从图像空间到不同特征空间的变换 T 各不相同,我们将在本章的后面几节中简要给出纹理和形状的特征空间以及变换 T 的简要描述。

5.3.2 基于四叉树的空间关系表达

空间关系表达的主要任务是指出具有某种特征的对象在图像中的位置关系。这个对象可以通过用户手工标定或系统自动分割产生的区域和区域的相对位置指出。在应用中主要要回答用户“在何处”的逻辑。

对大规模的图像集合进行手工标定区域是相当费时费力的,带有标定者的主观臆断,具有二义性,而且大多数的情况下也是不可能的,除非有足够的理由或专门的目的。所以空间关系的表达的核心工作便是区域的自动分割 (Segmentation)。同时,区域分割所产生的区域边界也可以作为图像的形状特征进行表达。区域分割的依据是区域内颜色和纹理的相似程度。

自动的分割并不是一件简单的事情。它需要一个唯一的数学算子,既能辨别大量的特征,又能精确地界定特征边界。这已经被证明是互相冲突的。因此,我们采用的基于四叉树的空间分割方法中,分割的边界是通过块状区域粗略估计的。这样做是为了换取在鉴别特征、求证特征相似性时更高的有效性。文献[20]给出了基于纹理特征的四叉树分割索引图像数据库的具体论述。

一个典型的四叉树自顶向下的分解过程,是将单个父节点分解为四个子节点,并递归地进行逐步分解,直到某个最小的限度,最后形成一棵正则的四叉树。这里我们根据给定条件进行图像的四叉树空间分割,同时在分割完成后增加了一个后处理模块用以拼接相邻的足够相似的子节点。这种拼接可以缩小四叉树的平均深度和广度以换取效率。这种对分割后的结果分块进行组合拼接策略可以加以扩展,以拼接所有相似相邻的块以获得任意属性区域的形状,以满足某些基于形状的应用,对此技术在此不作进一步的讨论。

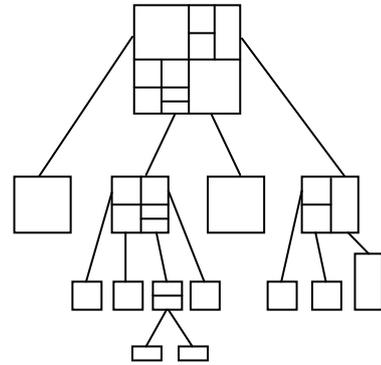


图 5-1 基于四叉树的分割

为了适应拼接的需要,我们需要修改四叉树的结构,允许四叉树的父节点可以拥有 4 个或少于 4 个的子节点。这样除了叶子节点外,每个节点可能会有 2 个, 3 个或 4 个子节点 (图 5-1)。

使用自顶向下的方法,我们将在每个节点对应的空间分块上进行特征抽取,根据抽取的结果,分解出所有的,或部分的子节点,当然也可能确认当前节点为叶子节点。那么,对每个空间分块,就需要进行特征抽取、父子节点间的距离计算、与作为子节点拼接条件的距离阈值进行比较等一系列的工作。

§ 5.4 基于颜色特征的检索

颜色特征是图像最直观、最明显的特征,一般用直方图描述。直方图的横轴表示颜色等级,纵轴表示在某一个颜色等级上具有该颜色的像素在整幅图像中所占的比例。根据图像的不同特点,可以采用不同的方法对图像进行预处理,然后用直方图进行匹配,以满足不同的检索要求。

认知科学及视觉心理学证明,人类不能像计算机显示器那样只使用 RGB 成份感知颜色,因此,选择一个适合于人类视觉特征的颜色空间可以改善检索效果。

实验证明,HSV 模型是一种适合人眼分辨的模型,在基于内容检索中用这种模型更适合用户的肉眼判断。这种颜色模型把彩色信号表示为三种属性:色调(Hue)、饱和度(Saturation)和亮度(Value)。色调 H 表示从一个物体反射过来或透过物体的光波长,亮度 V 是颜色的明暗程度,色度或饱和度 S 指颜色的深浅。例如,同样是红色,会因浓度不同而分为深红和浅红。一般一幅图像的颜色非常多,尤其是真彩色图像,因此,直方图矢量的维数也会非常多。如果对 HSV 空间进行适当的量化后再计算直方图,则计算量要少得多。所以,将 H、S、V 三个分量按照人的感知进行非等间隔量化,然后通过数学公式把 HSV 三维空间中的特征矢量转换为二维空间中的特征矢量。转换工作会带来误差,但对检索结果影响不大。

由于所采用的彩色空间与人的感知是一致的,因此,用转换后的直方图计算的差值对应于感知上的差别,这样可以比较明显地区分颜色上不相似的图像。

主色调查询主要使用颜色特征。主色调能够代表一幅图像的基本概貌,如蓝色主色调往往与大海或蓝天的图像相关,如果用户想要查找大海的照片,则可以指定蓝色作为主色调。但主色调仅仅反映了图像的大致情况,由于人的肉眼的分辨率有限,因此,可以选取两种以上颜色作为主色调(文献[4])。

有时,用户只想找具有某种颜色的目标,而对背景并不感兴趣。在这种情况下,必须将图像目标标识出来,即对图像进行分割,将背景滤掉。目前已有不少分割算法。在具体应用中,可以根据图像分割算法由计算机自动分割,也可以人工参与进行半自动分割。对分割后图像目标的颜色进行检索,可以满足用户基于色彩的模糊查询。

我们将在下一章中对基于颜色特征的检索就我们的工作展开论述,所以在此不再作进一步的展开。

§ 5.5 基于纹理特征的检索

纹理是图像中一个重要而又难以描述的特性。很多图像在局部区域内可能呈现出不规则性,而在整体上却表现出某种规律性。习惯上把图像中这种局部不规则而整体有规律的特性称之为纹理。从人类的感知经验出发,纹理特征主要有粗糙性、方向性和对比度(文献[1]),这也是用于检索的主要特征。

已有不少纹理分析方法,大致可分为两类。

(1) 统计方法

用于分析像木纹、沙地、草坪等细致而不规则的物体,并根据关于像素间灰度的统计性质对纹理规定出特征及特征与参数间的关系。

(2) 结构方法

适于像布料的印刷图案或砖瓦等一类元素组成的纹理及其排列比较规则的图案,然后根据纹理基元及其排列规则来描述纹理的结构及特征、特征与参数间的关系。

由于纹理难以描述,因此,对纹理的检索大都采用 QBE(Query By Example)方式。另外,

为缩小查找纹理的范围, 纹理颜色也作为一个检索特征。通过对纹理颜色的定性描述, 把检索空间缩小到某个颜色范围内, 然后再以 QBE 为基础, 调整粗糙度、方向性和对比度三个特征, 逐步逼近要检索的目标。

检索时首先将一些大致的图像纹理 (可采用 Brodatz 的纹理相簿, 大多数的纹理方面的研究都是针对 Brodatz 纹理集进行的 (文献[1]), 该集合给出了通常我们所见的各类纹理, 如图 5-2 所示) 以小图像的形式全部显示给用户, 一旦用户选中了其中某个与查询要求最接近的纹理形式, 系统则以查询表的形式让用户适当调整纹理特征, 如方向上“再偏左一点”, 粗糙度上“再细致一些”, 或对比度“再强一些”等, 通过将这些概念转换为参数数值进行调整, 并逐步返回越来越精确的结果。

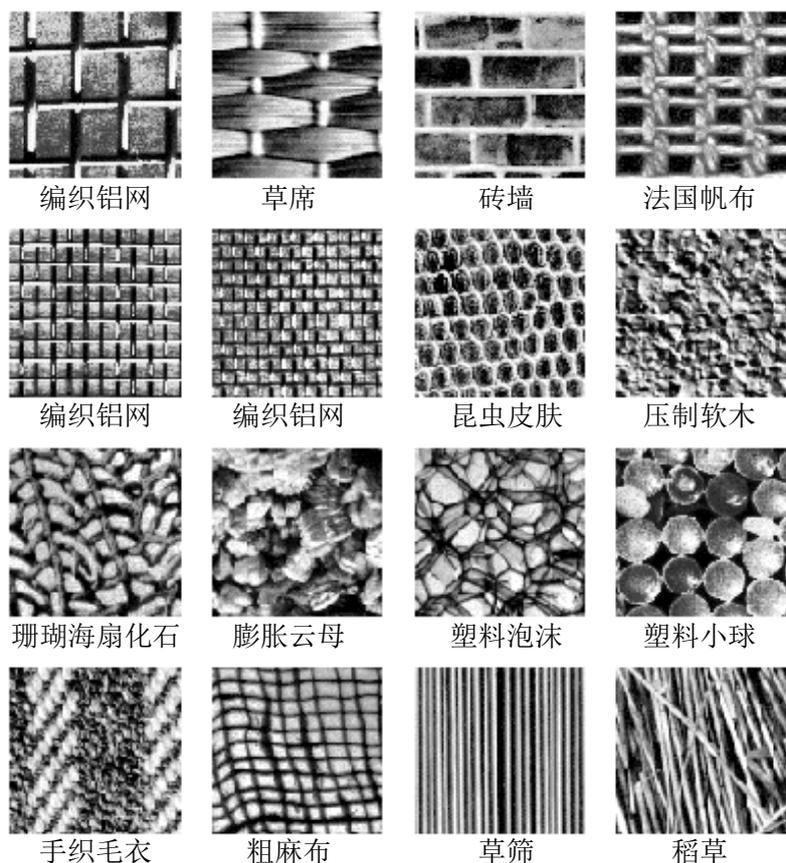


图 5-2 Brodatz 纹理集中的部分纹理

对于纹理的表达我们采用类似于颜色表达的多通道的表达方法。纹理图像首先通过小波滤波器组 FB (Filter Bank) 被分解为多个空间频率域的次能带 (Subband), 然后用纹理通道生成器 TCG (Texture Channel Generator) 生成所需要的多个通道。类似于颜色空间 (请参阅下一章), 我们通过纹理通道的变换 (T_i) 和量化 (Q_i) 方法, 以产生一个纹理元素的特征空间, 定义有效的纹理直方图 (h_i) 和二值纹理集 (s_i) 来表达纹理内容。

我们可以使用正交镜像滤波器 (QMF, Quadrature Mirror Filter) 组。QMF 小波滤波器组能在空间频域中产生倍频程 (Octave) 带宽的分割, 允许同时在空间域和空间频域获得较高的解析度, 而且, 使用小波组进行图像分析可以得到视觉机制的支持。

对图像中的每一个点, 通过 FB 和 TCG 在 k 个通道上分别生成了不同的能量值 (文献[11] 给出了 9 通道特征的完整论述)。这时, 纹理特征的表达已经类似于下一章中给出的关于颜色的在 HSV 空间中的三通道表示, 只不过对纹理我们面对的是 k 维的特征空间。

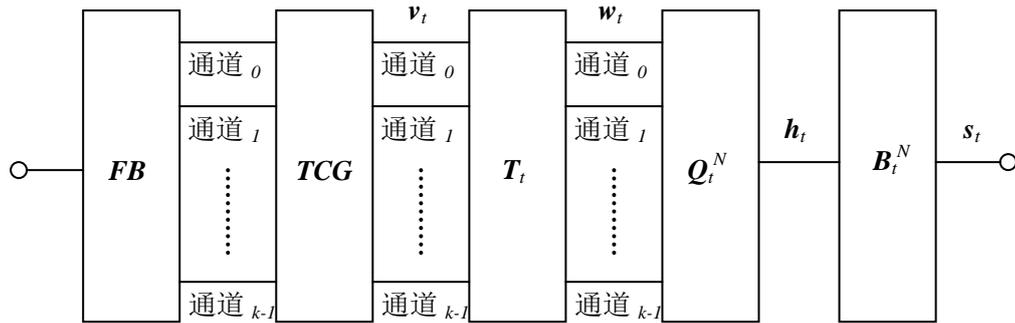


图 5-3 二值纹理特征集生成

大多数关于纹理特征的论著，都采用了灰度共生矩阵的表示方法，基于灰度共生矩阵的纹理分析算法如下：

灰度共生矩阵元素 p_{jk} 表示在与灰度为 j 的像素相距 δ 的像素中，灰度为 k 的像素之概率。这里， $\delta=(D_x, D_y)$ 。其中， δ 为位移矢量， D_x 为对当前点 X 方向的偏移量， D_y 为 Y 方向的偏移量。根据 p_{jk} 可分析出这种原色灰度变化的频繁程度。

灰度共生矩阵和我们所采用的滤波分解方法从本质上来说都属于一种统计方法。求出灰度共生矩阵后，可以计算出一组参数，用来定量描述纹理特性，其中能量、熵、惯性距、相关性、局部平稳性是比较常用的参数，但其中测量距离 δ 和测量方位 θ 的选择依赖于具体的纹理，很大程度上从而影响了它的应用（文献[1]）。

§ 5.6 基于形状特征的检索

形状特征是图像目标的一个显著特征，很多查询可能并不针对图像的颜色，因为同一物体可能有各种不同的颜色，但其形状总是相似的。如检索某辆汽车的图像，汽车可以是红的、绿的等，但形状决不会如飞机的外形。另外，对于图形来说，形状是它唯一重要的特征。

从图像中提取的目标边缘称为轮廓。基于形状或轮廓的检索是基于内容检索的一个重要方面，它能使用户通过勾勒图像的形状或轮廓，从图像库中检索出形状相似的图像。一个封闭的形状具有许多特征，如形状的拐点、重心、各阶矩，以及形状所包含的面积与周长比、长短轴比等。对于复杂的形状，还有孔洞数及各目标间的几何关系等。图形特征还包括其矩阵表示及矢量特征、骨架特征等。

基于形状特征的检索方法有两种：

(1) 分割图像经过边缘提取后，得到目标的轮廓线，针对这种轮廓线进行的形状特征检索。

(2) 直接针对图形寻找适当的矢量特征用于检索算法。处理这种结构化检索更为复杂，需做更多的预处理。

基于形状的检索更多地用于当用户粗略地画出一个轮廓进行检索的情况。这种轮廓可以是用户凭借脑子中的印象徒手画出来的，也可以是通过系统提供的基本绘图工具“拼凑”的。这两种情况都有一个特点，即提供的形状只是欲检索形状的粗略描述，它从大小、方向或整体结构上都可能与真正要查的图形有较大出入。因此，基于形状检索的难点在于寻找能够检索与大小、方向及扭曲伸缩无关的方法。

文献[6]给出了如下的一些形状特征量：

(1) 圆形度

$$\text{圆形度} = 4\pi \frac{(\text{面积})}{(\text{周长}) \times (\text{狭长度})}$$

(2) 狭长度。这是一个大于 1 的量, 能反映对象的狭长程度

$$\text{狭长度} = \frac{M_{\theta_{\max}}}{M_{\theta_{\min}}}$$

其中 M_{θ} 表示通过原点斜率为 θ 的直线 $y = x \tan(\theta)$ 周围的 $N \times M$ 图像 $f(i, j)$ 的惯性力矩:

$$M_{\theta} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (j \cos \theta - i \sin \theta)^2 f(i, j) = \mu_{02} \cos^2 \theta + \mu_{20} \sin^2 \theta - \mu_{11} \sin 2\theta$$

其 $(p+q)$ 次中心力矩 $u_{pq} = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (i - i_g)^p (j - j_g)^q f(i, j)$, (i_g, j_g) 为对象的重心

由 $dM_{\theta} / d\theta = 0$ 可求得 $M_{\theta_{\max}}$ 和 $M_{\theta_{\min}}$.

(3) 惯性主轴方向

$$\text{惯性主轴方向} \theta_0 = \frac{1}{2} \tan^{-1} \frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \pm \frac{n\pi}{2}$$

(4) 二阶矩常量

$$M_1 = \mu_{20} + \mu_{02}$$

(5) 多边形的顶点数

(6) 内角平均值, 是内角特征的一个总体描述, 能用来区分诸如三角形和四边形的形状。

$$a_v = \sum_{i=0}^{N-1} \frac{a_i}{N}$$

(7) 内角标准方差, 是对内角分布情况的一个总体描述, 多边形越规则, 其值越小。

$$\delta = \sqrt{\sum_{i=0}^{N-1} (a_i - a_v)^2}$$

(8) 多边形的凹点数, 可以用来区分凸多边形和凹多边形。

(9) 内角直方图, 可以用内角概率分布 P_i 表达, 其中 $i=1, 2, 3, 4$, 它们分别对应的内角区间分别是 $[0^\circ, 90^\circ)$, $[90^\circ, 180^\circ)$, $[180^\circ, 270^\circ)$, $[270^\circ, 360^\circ)$ 。

上述这些特征中, 绝大多数度与对象的方向、位置、尺寸无关。我们认为适合作特征空间方向的有圆形度、狭长度、二阶矩常量、内角平均值、内角标准方差, 多边形的凹点数、内角概率分布。

对图像中的每个形状, 计算其特征向量, 图像中落到该形状中的点被赋予此特征向量。若有相交的形状, 相交部分的点重复计数。对特征空间进行量化 Q_s 时, 对非封闭的方向根据统计规律进行非等间距量化, 然后得到特征直方图 h_s 。进一步的工作就是压缩至一维的二值形状集 s_s 。

§ 5.7 检索效果评价方法

基于内容的图像检索方法很多, 在具体应用中只有采用恰当的检索算法才能得到满意的结果。因此, 需要对这些算法进行评估, 比较其优劣。检索算法评价方法能够在相同的条件下找出最佳算法。在目前已有一些评价准则中, 主要有以下两种:

1、以查全率和查准率为基准的方法

查全率和查准率一般定义为:

查全率=检索到的相关图像数目/所有相关图像数目;

查准率=检索到的相关图像的数目/已检索出的图像数目。

为了评价算法的优劣,用户可以选定含有特定目标的图像作为一组相关图像,然后根据返回的结果计算查全率和查准率。查全率和查准率越高,说明该检索算法越有效。

2、排序评价方法

如果图像匹配采用相似性比较,那么,根据不同的相似性尺度,返回的图像数目可能不同,此时采用该方法。方法是:

(1) 选定一定数目的含有特定目标的相关图像,如水池、草坪、人与人会谈等。

(2) 固定返回图像的数目,把返回的图像按相似性大小排序。

(3) 计算有关返回图像的排序评价指标。比如计算检索到的相关图像在所有返回图像中的序号平均值。在理想情况下,所有相关图像都排在最前面,因此,平均值越小,表明检索算法越好。还可以计算丢失的相关图像占有所有返回图像的比例,这个值越小,表明成功率越高。

§ 5.8 小结

本章中我们对 CBIR 技术进行了一个简要的综述,分析了 CBIR 的技术特点和它的一般工作流程。

我们在这方面的主要工作是提出了一个通过直方图和二值集方法解决 CBIR 问题的有效的技术路线,使得我们能以一用一种一致的、紧凑的、自然的、完整的方法来看待和表达图像的颜色、纹理和形状特征,通过四叉树特征分解有效地表达图像中对象的空间关系。

在这个思路下,在本章中我们还就颜色、纹理和形状特征的表达作了介绍。其中主要对纹理和形状特征作了有限的展开。下一章对基于颜色特征的图像检索的充分展开将使得这一技术路线更加清晰。

第六章 基于颜色特征的图像检索

我们将给出一种用于图象数据库检索的视觉内容特征表达的统一框架。主要的目标是下列工作提供基础:

- (1) 度量图象视觉特征的相似性;
- (2) 存储和索引视觉特征信息;
- (3) 抽取图象中明显的区域边界。

这些目标将通过用直方图表示的内容特征来完成,直方图给出了图象或图象区域的特征分布信息。直方图定义在特征空间中,通过恰当地设计特征空间,直方图充分地表达了颜色和纹理的视觉度量,并满足上述目标的要求。

本章首先就颜色表达展开论述,我们将产生和评价颜色的特征空间,然后用直方图和二值集表达颜色特征。通过对图象数据的变换 T 和量化 Q 处理,我们给出这种视觉特征所对应的特征空间。工作的主要任务是产生颜色的一致、完整、紧凑和自然的特征空间而设计出有效的变换和量化对 (T_c, Q_c) 。

随后我们将讨论颜色特征空间的压缩、距离的计算和相似性度量的问题并给出具体的实现算法。

在本章中我们将看到基于直方图和二值集解决基于内容检索问题的总体思路。本章所论述的方法的基本思想同样适用于基于纹理和形状特征的检索。

§ 6.1 特征表达

对颜色这种视觉特征来说,一个图象点 (Image Points) 可以被初始化为多维的特征点 (Feature Points): 一个颜色点 (Color Point) 可以通过三个基本色定义。我们将给出具体的方法,使得特征点集通过变换和量化产生由 194 个颜色元素构成的特征空间。

使用这些特征元素,我们将应用两种方法来表达颜色:直方图和二值集。直方图 (Histogram) 是特征元素的一种分布,而二值集 (Binary Set) 是特征元素的一种逼近和压缩。

§ 6.2 一些相关工作和比较

颜色直方图用于图象检索的关键在于颜色特征空间 (T_c, Q_c) 的选取和直方图距离 D_i 的度量。目前对哪一种颜色特征空间更适合基于颜色直方图的检索尚无一致的意见。问题在于虽然不存在一种普遍适用的颜色空间,彩色视觉仍然是一种非常重要的视觉要素。因此,在实践中我们尝试使用了多种的颜色空间。

目前有很多学者就基于颜色的图像检索开展了一些有意义的工作,如文献[4]的作者提出了一种基于主色调的图像检索方法,该方法强调颜色的视觉特性,该方法的主要思想已经在本文所述方法的 HSV 颜色空间的量化中得到反映,但文献[4]没有考虑在大规模的图像集合上实现的空间和时间效率,其实验结果是在 600 幅的小规模集合上作出的,主要追求检索效果。

文献[3]在基于色彩的图像数据检索方面的研究上则更注重效率与效果的均衡,在检索方法的实现思路上与本文论述的方法很接近,比如二值集逼近、多分辨率、多相似粒度的检索等方法。但本文的工作更强调与领域和实践结合。本文在进行直方图的相似性比较时,直

接采用二值集的逼近比较,以使得的特征预定义在数据库的存储上成为可能,并得到较高的检索效率。文献[3]之所以能够采用颜色直方图直接进行相似性度量,是因为采用了所谓的“降低直方图颜色维数”的方法,如将一幅真彩色的图像转换到 64 或 256 色后,进行相似度的运算,虽然能够提高检索效率,但没有对颜色的视觉特性予以充分的重视,在实际的数据库存储上也并不优于本文的方法(本文方法中的 194 个颜色元素构成的二值特征空间只需要 25 个字节进行描述),并且在医学影像的应用中需要尽可能多地保留视觉信息。

本章将就颜色空间的选取和评价提供一个工作构架,并在此基础上给出基于颜色直方图的多相似性检索方法。我们将首先讨论颜色的基本性质,以及彩色视觉的色度学。

§ 6.3 彩色视觉的色度学和彩色感觉

可见光 ($\lambda \in [380\text{nm}, 780\text{nm}]$) 的电磁辐射 ($P(\lambda)$) 被人们感觉为颜色或色光。实验证明,彩色视觉是通过三种独立的感受器形成的,它们分别对波长为 $\lambda_r=700\text{nm}$, $\lambda_g=546.1\text{nm}$, $\lambda_b=435.8\text{nm}$ 的红绿蓝三色最为敏感。对每个基色感受器 $k \in (r:g:b)$, 设其响应函数为 $C_k(\lambda)$, 对任意的可见光的颜色可以表示为 $C_k(\lambda)$ 的线性叠加(文献[2])。

将 $C_k(\lambda)$ 归一化后,白光 $W(\lambda)$ 可以表示为:

$$W(\lambda) = \overline{C_r}(\lambda) + \overline{C_g}(\lambda) + \overline{C_b}(\lambda)$$

$P(\lambda)$ 是对三基色 (R, G, B) 的响应:

$$P(\lambda) = RC_r(\lambda) + GC_g(\lambda) + BC_b(\lambda) \dots \dots \dots (1)$$

这样,任何颜色均可用三基色 (R, G, B) 的线性组合来表示。

彩色感觉是一个复杂的过程,我们有必要做一些简化的假定:我们假设颜色视觉是一种针对像素的感觉,而不受到周围颜色环境的影响。而且我们也假设诸如环境光照、观察者的适应性、观察距离、图像的显示质量等观察条件是可以忽略的。一般说来在不同的图像检索的应用中,这些参数是很难控制的。

§ 6.4 颜色空间的变换和量化

在采用 RGB 值表示的颜色空间中,所有的颜色皆被表示为 3-D 的 RGB 空间中的一个向量。我们把这个空间作为我们表达图像颜色特征的起点。我们研究不同的变换 T_c 和相应的量化 Q_c , 以得到一些相互分隔的空间,而空间都应具有一致性、完整性、紧凑性和自然性。这些空间都允许用颜色直方图和二值集表示图像颜色内容。

属性1 一致性 颜色之间的近似度量应表达视觉感知上的相似性。

后面我们将看到,基于颜色的图像比较利用了一种颜色特征的近似度量方法。这种比较要求颜色近似程度应易于计算。有一点可以做到的是,至少我们所设计的空间变换应使得颜色近似程度不是颜色在空间中位置的函数。变换 T_c 对颜色空间的一致性起着决定性的作用,同时也决定了相应的量化 Q_c , 以及颜色空间的完整性与紧凑性。

属性2 完整性 颜色空间应包括所有的可感知的不同色彩。

在颜色空间中包括所有可感知的不同颜色是必要的。视觉上的完整性并不保证数学上的完整性,但数学上的完整性能够保证视觉完整。一般说来,如果变换 T_c 是可逆的,未量化的颜色空间将是完整的。在这种情况下,量化 Q_c 将决定完整性。

属性3 紧凑性 颜色空间中每种颜色应在视觉上区别于其他颜色。

为了限制图像颜色特征表达的维数,或者说颜色的数目,颜色空间应是非冗余的或者

说是紧凑的。在数学上,当从 RGB 空间进行一对一或多对一映射时,非冗余就能够得到保证。但在视觉上,非冗余要求不存在近似到让人感觉是同一种颜色的两种颜色。一般说来,量化 Q_c 的选择将使得我们必须完整性和紧凑性之间作出折衷。

属性4 自然性 颜色空间应能方便于将颜色自然分解为三个基本的视觉特性(亮度、色调、饱和度)

将颜色分解为亮度、色度、饱和度三个元素是一种人类知觉。在颜色空间中定位的难易程度直接影响用户构造基于颜色查询的能力。

亮度(Brightness)是指彩色光所引起的人眼对明暗程度的感觉。显然亮度和照射光的强度有关。

色调(Hue)是指光的颜色。红、橙、黄、绿……都表示不同的色调。改变色光的光谱成分,就会引起色调的变化。

饱和度(Saturation)是指颜色的深浅程度,如深红、淡红等。

色调和饱和度有合称为色度(Chrominance),它既表示色光的颜色类别,又能表示颜色的深浅程度。

我们设计变换 T_c 和量化 Q_c 以产生一个能够达到这些目标的颜色空间。要指出的是,这些条件和目标是与应用相关的。例如,医学和卫星图像需要和彩色肖像照片不同的颜色空间,以在颜色特征的表达上保证一致性、完整性,其次才是紧凑性。

首先,我们将一个颜色点(Color Point)定义为 RGB 空间中的一个向量:

定义1 颜色点 v_c 一个颜色点是 RGB 颜色空间中的一个向量 $v_c = (r, g, b)$ 。

RGB 颜色空间的变换 T_c 和量化 Q_c 将对向量 v_c 进行改造和分组,以创建颜色的有限集。

6.4.1 颜色变换

颜色变换 T_c 是对向量 $v_c \in RGB$ 的一种操作,以生成变换后的向量 w_c 。简单的颜色变换都是线性的。例如,从 RGB 可以通过线性变换得到 YIQ (NTSC 彩色电视标准), YUV (PAL 和 SECAM 彩色电视标准), $YCrCb$ (JPEG 数字图像标准和 MPEG 数字视频标准) 和 OPP (Opponent Color Space) 等。而其他的颜色空间,如 HSV , $CIE1976$ 都是由非线性变换产生的。总的说来颜色变换 T_c 的作用是产生一个新的颜色空间。

由于 RGB 颜色空间是连续的,或者有可能已用大量的值离散化(如数字照片常常对每个像素在 RGB 颜色空间中的每个分量上用 256 个灰度级进行离散化,这样颜色空间中将有 1.67M 的颜色点),颜色空间必须进行量化,或分割为较少数量的颜色。

6.4.2 颜色量化

由于颜色空间是多维的,对颜色空间的分割将描述为空间中向量的量化。一般地,一个 K 维和 M 规模的量化器,是指对 K 维空间中的向量到包含 M 个元素的有限集 C 的一种映射,这样,量化器 Q_c 可以定义为:

$$Q_c: R^k \rightarrow C$$

其中 $C = \{y_0, y_1, \dots, y_{m-1}\}$ 且 $y_m \in R^k$ ($m=0, 1, \dots, m-1$)。一般地,集合 C 称为译码本,并且规模为 M 。对颜色空间的量化来说, $k=3$,译码本中的每个条目对应于一个颜色点。

所以,译码本 C 就能表示全部颜色或其中的一部分。在这里,一个颜色,对应于颜色空间的一个量化分割 R_m ,是对颜色视觉单元的一个近似。它区别于上面提到的变换后的颜色点,后者是连续空间中的一个精确向量。与 M 相关的点向量量化器将 R^k 分割为 M 个部分,每个分割 R_m 包含所有被编码为 y_m 的所有颜色点 w_c :

$$R_m = \{w_c / w_c \in R^k, Q_c(w_c) = y_m\}$$

从分割的定义我们可以看到所有的分割将覆盖 R^k 且互不重叠:

$$\bigcup_m R_m = R^k \quad \text{并且} \quad R_m \cap R_n = \phi \quad \forall m \neq n$$

这样分割将是对 R^k 的完整分割。

定义2 颜色 (索引为 m) 一个“颜色”是对应于分割 R_m 的一个色彩单元, 由一组颜色点 $\{v_c\}_m$ 经过变换 T_c 和向量量化 Q_c 得出并被赋予同样的编码 $y_m = Q_c(T_c v_c)$.

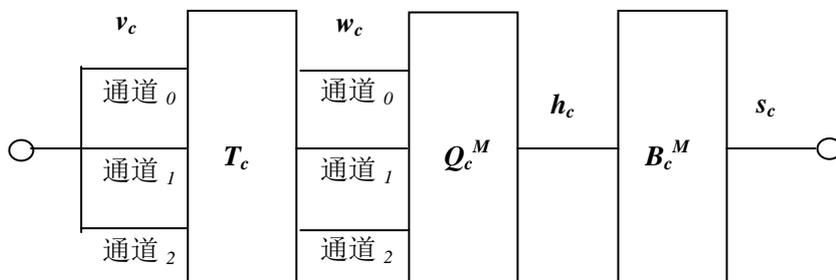


图 6-1 变换 T_c 和量化 Q_c 产生颜色特征空间, 用以定义颜色直方图 h_c 和二值集 s_c

§ 6.5 颜色空间

下面我们将给出几种从 RGB 颜色空间产生新颜色空间的变换 T_c , 并分别做出评价。以此为起点, 通过对 RGB 颜色空间进行变换, 设计出合适的颜色特征空间, 并给出颜色直方图和二值颜色集的定义。

6.5.1 RGB 颜色空间

RGB 颜色空间是首先可选的颜色空间。但是, RGB 颜色空间与视觉是不一致的。这样, RGB 颜色空间中颜色的接近并不能说明颜色的相似性。由于 RGB 颜色空间是完整的, 其他完整的颜色空间可以通过对 RGB 颜色空间作变换 T_c 得到, 而且变换也不必一定是线性的。

6.5.2 线性变换颜色空间

有一些颜色变换广泛地应用于表示、发送、广播彩色图像或视频信号。这些变换大都可通过简单的矩阵乘法进行计算。

6.5.2.1 OPP 颜色空间

OPP 颜色空间 (Opponent Color Space) 是对 RGB 颜色空间作线性变换得到。有证据表明, 人类彩色视觉使用一种反颜色模型, 将 R 、 G 、 B 锥体细胞的响应重组为两组反颜色通道, 以形成白昼视觉。从 RGB 到 OPP 的变换可以表示为:

$$w_c = T_c^{OPP} v_c$$

其中,

$$T_c^{OPP} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & 2 \\ 1 & -2 & 1 \end{bmatrix} \dots\dots\dots(2)$$

OPP 颜色空间的主要优点是能够很容易通过 T_c^{OPP} 得到，缺点是缺乏一致性和自然性。OPP 颜色空间中的颜色距离并不能很好地度量颜色的相异性。亮度 (Luminance) 是 OPP 通道之一，而另外的两个色度 (Chrominance) 通道并不对应与色调和饱和度，而只是简单地用蓝色对应黄色、红色对应绿色。

6.5.2.2 线性颜色变换标准及 YIQ 颜色变换

相当一部分彩色图像系统采用 YIQ, YUV 和 YCrCb 线性颜色变换。这些变换都是从 RGB 产生一个亮度 (Luminance) 通道和两个色度 (Chrominance) 通道，而且都是针对特定的显示设备专门设计的，如 YIQ 用于 NTSC 制式的彩色电视，YUV 用于 PAL 和 SECAM 制式的彩色电视，而 YCrCb 用于计算机彩色显示器。这些颜色空间都不具备一致性，在这些颜色变换空间中的颜色距离并不对应视觉相似性。

YIQ 颜色空间是一种用于 NTSC 制式彩色电视的颜色标准。从 RGB 到 YIQ 的变换可以表示为：

$$w_c = T_c^{YIQ} v_c$$

其中，

$$T_c^{YIQ} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \dots\dots\dots(3)$$

6.5.2.3 YUV 颜色变换

YUV 颜色空间是一种用于 PAL 和 SECAM 制式彩色电视的颜色标准。从 RGB 到 YUV 的变换可以表示为：

$$w_c = T_c^{YUV} v_c$$

其中，

$$T_c^{YUV} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & -0.436 \\ 0.515 & -0.515 & -0.100 \end{bmatrix} \dots\dots\dots(4)$$

6.5.2.4 YCrCb 颜色变换

YCrCb 颜色空间是一种用于 JPEG 数字图像的颜色标准。从 RGB 到 YCrCb 的变换可以表示为：

$$w_c = T_c^{YCrCb} v_c$$

其中，

$$T_c^{YCrCb} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5000 & -0.4187 & -0.0813 \\ -0.1587 & -0.3313 & 0.5000 \end{bmatrix} \dots\dots\dots(5)$$

尽管线性的颜色变换相对来说比较简单，但他们被能够不能产生具备一致性和自然性的颜色空间。下面我们会看到，Munsell 颜色表系统虽然相对复杂，却是针对自然性、紧凑性和完整性而定义的。

6.5.3 Munsell 颜色表系统

Munsell 颜色表系统 (Munsell Color Order System) 是一种颜色标记系统，它按照颜色的自然属性组织颜色。Munsell 在其著作《Book of Color》中给出了 1200 个颜色块(color chips) 的示例，并给出了相应的色调 (Hue)、饱和度 (Saturation) 和色品 (chroma)。我们可以认为，每个 Munsell 颜色块是通过一些对 RGB 彩色空间进行的变换 T_c 和量化 Q_c 得到的，这些颜色块通过良好的组织使其每个单元间隔与视觉系统的区别能力相等价。

Munsell 颜色表系统的主要优点是它给出了能够很好地表达视觉相似性的有限颜色集，用以取代三维的视觉颜色空间。但颜色表系统有一个致命的缺陷，就是无法确切地给出对 RGB 彩色空间进行的确定的变换 T_c 和量化 Q_c ，来产生这个有限颜色集。尽管 Miyahara 给出了一种称为 MTM (Mathematical Transform to Munsell) 的变换，在其图像数据研究中用于 RGB 到 Munsell HVC 的转换，但仍然不存在一种从 RGB 颜色空间中的颜色点到 Munsell 颜色块的简单映射方法。Munsell 颜色表系统是紧凑的 (1200 个视觉上相区别的颜色块) 和视觉完整的，但是，并不满足一致性的要求。Munsell 颜色表系统并不是针对不相邻的颜色块之间的相似度的度量而设计的。

6.5.4 CIE 颜色空间

针对颜色空间一致性的目标，CIE (国际照明委员会) 于 1976 年定义了 CIE1976 (Luv) 和 CIE1976 (Lab) 两种颜色空间。对 RGB 颜色空间首先通过线性变换得到 XYZ 颜色空间，然后通过一个或两个非线性变换得到 CIE1976 (Luv) 或 CIE1976 (Lab) 颜色空间。RGB 和 XYZ 颜色空间之间的线性变换可以表示为：

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \dots\dots\dots(6)$$

从 XYZ 颜色空间到 CIE1976 (Luv) 或 CIE1976 (Lab) 颜色空间的非线性变换，是根据与目标白色刺激的三基色 (X_n, Y_n, Z_n) 的关系定义的，这样，亮度 (Lightness) 可以由下式给出：

$$L = 116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{8}} - 16 \dots\dots\dots(7)$$

而 u, v 和 a, b 的值可以分别根据下面的一些公式计算：

$$\begin{aligned} u &= 13L(u' - u'_n) \dots\dots\dots(8) \\ v &= 13L(v' - v'_n) \end{aligned}$$

其中

$$\begin{aligned} u' &= \frac{4X}{X + 18Y + 8Z} & u'_n &= \frac{4X_n}{X_n + 18Y_n + 8Z_n} \\ v' &= \frac{4Y}{X + 18Y + 8Z} & v'_n &= \frac{4Y_n}{X_n + 18Y_n + 8Z_n} \end{aligned}$$

两个颜色刺激之间的颜色距离为：

$$\Delta L_{uv} = \sqrt{(\Delta L)^2 + (\Delta u)^2 + (\Delta v)^2} \dots\dots\dots(9)$$

对 CIE1976 (Lab):

$$a = 500 \left[\left(\frac{X}{X_n} \right)^{\frac{1}{8}} - \left(\frac{Y}{Y_n} \right)^{\frac{1}{8}} \right]$$

$$b = 200 \left[\left(\frac{Y}{Y_n} \right)^{\frac{1}{8}} - \left(\frac{Z}{Z_n} \right)^{\frac{1}{8}} \right] \dots\dots\dots(10)$$

此时两个颜色刺激之间的颜色距离为:

$$\Delta L_{ab} = \sqrt{(\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2} \dots\dots\dots(11)$$

CIE 颜色空间对重要颜色的表达是比较一致的,三个颜色分量非常好地刻画了颜色的色调 (Hue)、亮度 (Lightness) 和饱和度 (Saturation)。但是, CIE 颜色空间需要对 RGB 颜色空间进行正的或逆的非线性变换,具体应用时是不方便的,且需要一定的运算量。

6.5.5 HSV 颜色空间

HSV 用色调、亮度和饱和度表达颜色,从 RGB 到 HSV 的变换也是非线性的,但是更容易进行逆变换。HSV 颜色空间具备自然性,而且具有相当的视觉一致性。因此,我们可以定义一个 HSV 量化 Q_c ,以产生同时具备紧凑性和完整性的颜色集合。

6.5.5.1 HSV 变换

设 v_c 是 RGB 颜色空间中的一个颜色点, w_c 是 HSV 颜色空间中的一个颜色点,从 RGB 到 HSV 的变换 T_c 使得 $w_c = T_c(v_c)$, 则变换 T_c 可以定义如下:

变换 1 RGB \Rightarrow HSV 对 $r, g, b \in [0 \dots 1]$, T_c 将根据下面的等式计算对应的 $h, s, v \in [0 \dots 1]$:

$$v = \max(r, g, b) \quad s = \frac{v - \min(r, g, b)}{v}$$

$$\text{定义 } r' = \frac{v - r}{v - \min(r, g, b)} \quad g' = \frac{v - g}{v - \min(r, g, b)} \quad b' = \frac{v - b}{v - \min(r, g, b)}$$

$$h' = \begin{cases} 5 + b' & \text{若 } r = \max(r, g, b) \text{ 且 } g = \min(r, g, b) \\ 1 - g' & \text{若 } r = \max(r, g, b) \text{ 且 } g \neq \min(r, g, b) \\ 1 + r' & \text{若 } g = \max(r, g, b) \text{ 且 } b = \min(r, g, b) \\ 3 - b' & \text{若 } g = \max(r, g, b) \text{ 且 } b \neq \min(r, g, b) \\ 3 + g' & \text{若 } b = \max(r, g, b) \text{ 且 } r = \min(r, g, b) \\ 5 - r' & \text{其他} \end{cases}$$

$$h = h' / 6$$

变换 1 HSV \Rightarrow RGB 对 $h, s, v \in [0 \dots 1]$, T_c^{-1} 将根据下面的等式计算对应的 $r, g, b \in [0 \dots 1]$:

$$\alpha = 6h - \text{round}(6h)$$

$$\omega_1 = (1-s)*v \quad \omega_2 = (1-(s*\alpha))*v \quad \omega_3 = (1-(s*(1-\alpha)))*v$$

$$r = \begin{cases} v & \text{若 } \alpha = 0 \text{ 或 } \alpha = 5 \\ \omega_1 & \text{若 } \alpha = 2 \text{ 或 } \alpha = 3 \\ \omega_2 & \text{若 } \alpha = 1 \\ \omega_3 & \text{若 } \alpha = 4 \end{cases}$$

$$g = \begin{cases} v & \text{若 } \alpha = 1 \text{ 或 } \alpha = 2 \\ \omega_1 & \text{若 } \alpha = 4 \text{ 或 } \alpha = 5 \\ \omega_2 & \text{若 } \alpha = 3 \\ \omega_3 & \text{若 } \alpha = 0 \end{cases}$$

$$b = \begin{cases} v & \text{若 } \alpha = 3 \text{ 或 } \alpha = 4 \\ \omega_1 & \text{若 } \alpha = 0 \text{ 或 } \alpha = 1 \\ \omega_2 & \text{若 } \alpha = 5 \\ \omega_3 & \text{若 } \alpha = 2 \end{cases}$$

6.5.5.2 HSV量化

我们设计了一个 HSV 量化器 Q_c ，从 HSV 颜色空间产生一个包含 194 个元素的紧凑颜色集。我们用柱面坐标系来描述 HSV 颜色空间（如图 2 所示）。柱轴方向表示亮度（Value），到柱轴的距离表示饱和度（Saturation），绕柱轴的角度表示色调（Hue）。

在色度盘的圆周（色调）上，红、绿、蓝三原色互相间隔 120 度。将圆周用 20 度等分就已经足够用以区分不同的色调，这样三原色和黄、品红、青色都至少有三个等分部分加以细分。饱和度和亮度均被分为 3 个等级，已经足够满足视觉在这两个通道上的要求。考虑到在医学影像中有较多的灰度图像，再加上 32 种灰颜色，HSV 量化器 Q_c 对 HSV 颜色空间给出了 194 种相互区分的颜色。

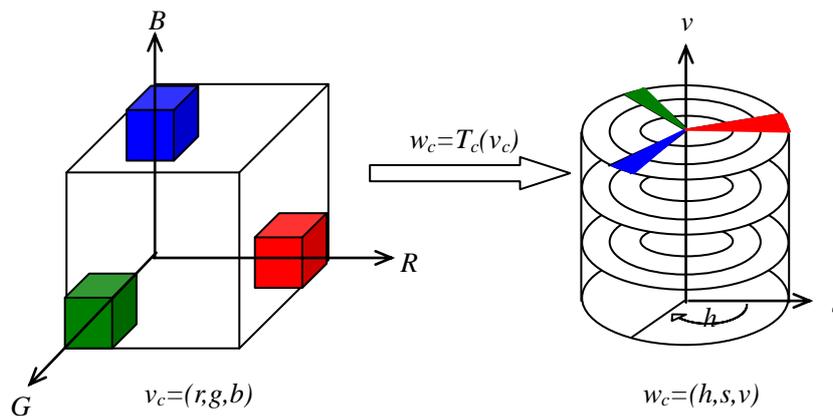


图 6-2 从 RGB 到 HSV 的变换 T_c 和量化 Q_c 给出了 18 种色调、3 种饱和度、3 种纯度，加上 32 个灰度色，共 194 种颜色

6.5.5.3 HSV 颜色相似度

由于 HSV 颜色空间在视觉上是近乎一致的，所以，两个 HSV 颜色的相似度可以近似地用它们在 HSV 颜色空间中的距离来度量。一般地，设有两个索引值分别为 $m_i = (h_i, v_i, s_i)$

和 $m_j = (h_j, v_j, s_j)$ 的颜色，它们的近似度可以定义为相应的颜色点在 *HSV* 颜色空间的柱面坐标系中的距离：

$$a_{i,j} = 1 - \frac{1}{\sqrt{6}} \sqrt{(v_i - v_j)^2 + (s_i \cos(h_i) - s_j \cos(h_j))^2 + (s_i \sin(h_i) - s_j \sin(h_j))^2} \dots\dots\dots(12)$$

由于对相似度进行了归一化，对索引值分别为 i 和 j 的两个颜色，若 $a_{i,j} = 1$ ，则表示最为接近；若 $a_{i,j} = 0$ ，则表示在 *HSV* 颜色空间中的距离为最大，在视觉上差异也最大。

6.5.5.4 各类颜色变换的比较

表 6-1 对照比较了前面提到的各类颜色变换。其中，*OPP*, *YIQ*, *YUV* 和 *YcrCb* 颜色空间是从 *RGB* 颜色空间通过线性变换得到的。从 *RGB* 到 *HSV* 的变换是非线性的，也是可逆的，同时变换也是易于计算的。在可选的颜色空间中，只有量化的 *HSV* 颜色空间满足一致性、完整性、紧凑性和自然性。*Munsell* 和 *CIE* 颜色空间虽然满足自然性条件，但是 *Munsell* 颜色表系统不均衡，*CIE* 颜色空间不够紧凑。

	RGB	OPP/YIQ/YUV/YcrCb	Munsell	CIE	194-Color HSV
一致性	×	×	×	√	√
紧凑性	×	×	√	可能	√
完整性	√	√	√	√	√
自然性	×	×	√	√	√
变换		线性	非线性	线性+非线性	非线性

表 6-1 各类颜色变换的比较

§ 6.6 表达颜色特征

使用 194 色量化的 *HSV* 颜色空间，我们可以通过度量这 194 种颜色在图像中的分布来表达图像的颜色特征。

6.6.1 颜色直方图

图像中颜色的分布可以用直方图来表示。对于给定的图像 $I[x, y]$ ，有三个颜色通道， $I = (i_R, i_G, i_B)$ ，它的颜色直方图可以由下式给出：

$$h_c[m] = \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} \begin{cases} 1 & \text{若 } Q_c(T_c I[x, y]) = m \\ 0 & \text{否则} \end{cases}$$

其中， X , Y 分别为图像的宽度和高度。

定义 3 颜色直方图 h_c 颜色直方图定义了图像或图像中区域的颜色分布。

6.6.2 颜色直方图的性质

针对静止图像的检索，可使用的特征有颜色、纹理、形状等。颜色直方图在传统的图理解中未能受到足够的重视，然而相对于几何特征而言，颜色直方图具有与生俱来的旋转不变性（rotation-invariance）和缩放不变性（scale-invariance）。

图 6-3 显示了一张 CT 切片在其亮度通道中的旋转不变性和缩放不变性。当然不变性是有前提的，也就是说在旋转和缩放过程中没有信息的丢失和增加。由于图 6-3 中的图像都是 JPEG 格式压缩存储的，加上在缩放过程中采用了简单的统计方法，所以在旋转和缩放过程中有少量的信息丢失，3 张图片的直方图在细节上出现了少许的差异，但在总体的曲线形态和能量分布上是十分相似的。这也正符合了我们进行基于内容的图像检索的目的，是为了检索相似而不是相等的图片。

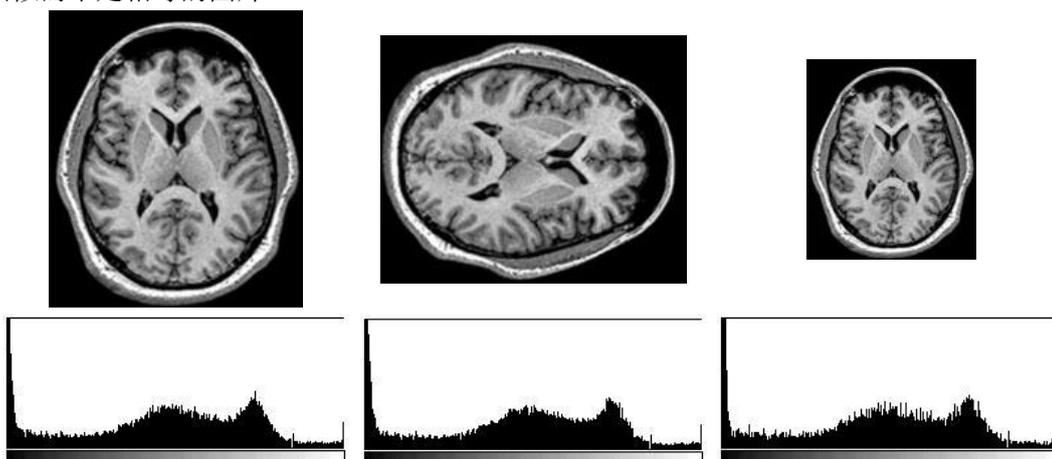


图 6-3 颜色直方图的旋转不变性和缩放不变性

6.6.3 二值颜色集

图像或图像区域有关颜色的内容可以更紧凑的二值颜色集来表示。二值颜色集可以从前面筛选出的 M 个颜色得到，图 1 描述了整个处理过程。设 B^M 为 M 维的二值空间，其中的每一维对应一个颜色，相应的索引为 m 。

定义 4 二值颜色集 二值颜色集是中的一个二值向量，定义了图像或图像区域中的颜色集合 $\{m\}$ 。

二值颜色集等价于阈值化的直方图，阈值使得每个直方图中的每一个值转化为两个可能的等级之一。例如，为颜色 m 给定阈值 τ_m ，相应的二值颜色集可以从下是得到（图 6-4）：

$$s_c[m] = \begin{cases} 1 & \text{若 } h_c[m] \geq \tau_m \dots\dots\dots(14) \\ 0 & \text{否则} \end{cases}$$

二值颜色集仅描述了有哪些颜色是大于阈值的，但在表达区域性的颜色是有效。当某个颜色不能在某个颜色区域中有效表示，比如，颜色低于阈值时，这个颜色将被忽略。这样，用二值颜色集的方法，只有在区域中较突出的颜色所包含的颜色内容能得到表达。

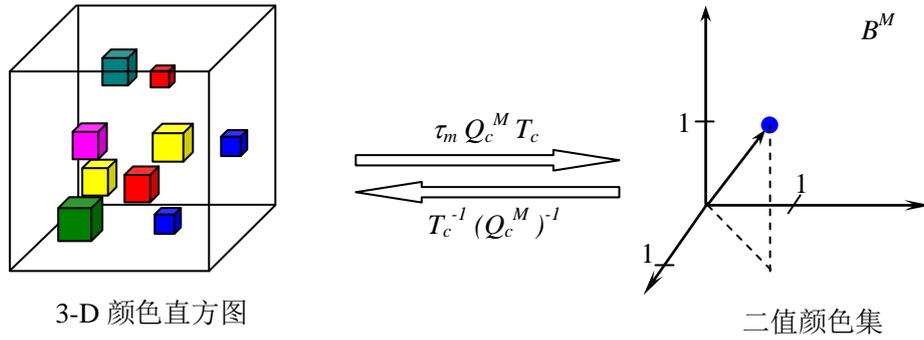


图 6-4 3-D 颜色直方图与二值颜色集之间的关系

6.6.4 颜色直方图压缩

由于我们的目标是有效地表达区域或图像的颜色信息，我们研究了典型的图像和区域的颜色直方图的特性。图 6-5 比较了手工抽取的图像区域和整幅图像的颜色直方图的能量紧缩情况，我们看到，图像中区域的颜色信息很大程度上由少量重要的颜色决定，而对整幅图像，颜色信息扩散到许多颜色。由此我们得到这样的做法：

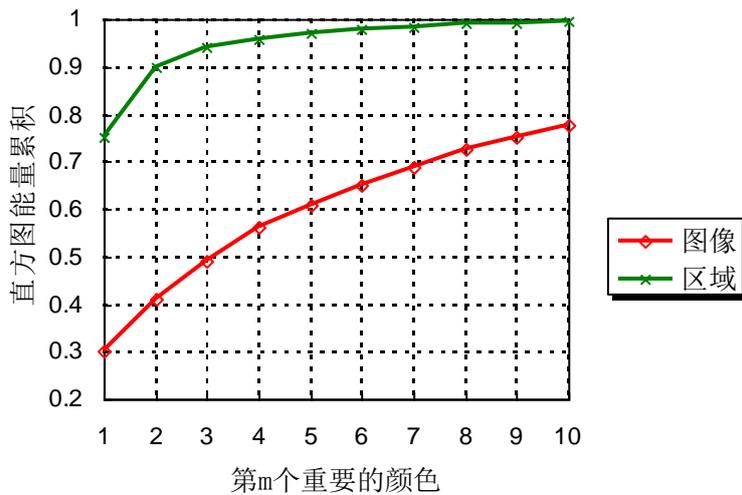


图 6-5 区域和图像的直方图能量累积情况比较

为了表达区域性的颜色，我们可以使用诸如二值颜色集的紧凑特征集。由于区域中只含有少量的颜色，我们假定目标图像中与手工描述的区域相对应的部分也只含有少量的颜色。在这个假定的前提下，我们研制了一个方法以自动地从图像中抽取图像区域（基于四叉树的特征分割方法）。

更进一步，我们希望能够开发出多功能的颜色直方图和二值颜色集，为此，我们也将给出一个从直方图到最相近的二值集的快速映射算法 BSFM（Binary Set Fast Mapping），我们用这个算法来提高用二值集绑定（BSB, Binary Set Bounding）的直方图检索的效率。

6.6.5 二值集快速映射 (BSFM)

这里我们将给出 BSFM(Binary Set Fast Mapping)算法,用以对任意的直方图求得最相近的二值集。主要的目标是寻找一个二值集 s_c , 使得其与给定直方图 h_c 的距离最小, 也就是说均方差 $e_c(h_c, s_c)$ 最小。这里,

$$e_c(h_c, s_c) = \sqrt{\sum_{m=0}^{M-1} \left(h_c[m] - \frac{s_c[m]}{|s_c|} \right)^2}$$

其中 $|s_c|$ 表示在二值集中非零元素的个数。该算法的步骤大值为:

- (1) 对 h_c 的所有元素按值从大到小排序。
- (2) 从最大的元素开始我们得到一个势为 1 ($|s_c|=1$) 的二值集, 并将二值集中的这个元素赋值为 1, 计算排序后的直方图和这个二值集的均方差。
- (3) 继续按从大到小的顺序依次处理直方图中的元素, 并将二值集中与此顺序对应位置的元素赋值为 1, 计算排序后的直方图和这个二值集的均方差。直到处理到某个元素时误差 e_c 达到最小。

通过这个选择过程, 我们最终能够得到原直方图均方差最小的, 也是最相近的二值集。

具体的处理过程叙述如下: 设映射 M_c 可以使得 h_c 重新排序,

$$g_c = M_c h_c, \text{ 且 } g_c[0] \geq g_c[1] \geq \dots \geq g_c[M-1].$$

我们也注意到, h_c 可以通过相应的逆映射从 g_c 得到:

$$h_c = M_c^{-1} g_c$$

设 r_c^n 是具有 M 个元素、势为 n ($|r_c^n|=n$) 的二值集, 其中,

$$r_c^n[m] = \begin{cases} 1 & m < n \\ 0 & m \geq n \end{cases}$$

且定义均方差 ε_c^n ,

$$\varepsilon_c^n(g_c, r_c^n) = \sqrt{\sum_{m=0}^{M-1} \left(g_c[m] - \frac{r_c^n[m]}{n} \right)^2}$$

我们可以很容易地确定使得 ε_c^n 最小的 n 。而且可以证明,

$$e_c(h_c, s_c) = \varepsilon_c^n(g_c, r_c^n)$$

则与直方图 h_c 最相近的二值集 s_c 可以由下式给出:

$$s_c = M_c^{-1} r_c^n$$

整个算法需要 $M \log M$ 次的操作用以对 h_c 进行排序, 最多 M 次的运算用以计算均方差和确定最小值。

§ 6.7 直方图的归一化与度量关系

直方图是离散分布, 所以可以通过 M 维空间的特征向量来表示。 M 表示直方图中颜色的种类数。我们可以给出一个一般的归一化公式:

$$h^{(r)} = \frac{h}{\sqrt[r]{\sum_{m=0}^{M-1} |h[m]|^r}} \quad (\text{其中 } r=1,2)$$

图 6-6 分别给出了 $r=1$ 和 $r=2$ 时的归一化的二维直方图在二维离散空间中的特征向量。

设 M 维空间中任意两个直方图 h_i, h_j 的差距为 $D(h_i, h_j)$ 。则直方图之间的可能的度量关

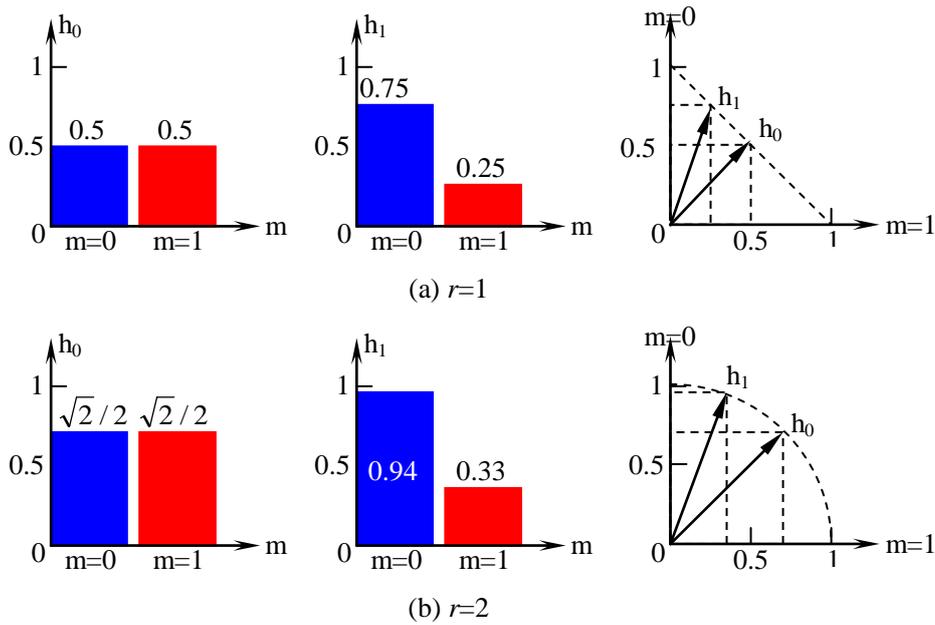


图 6-6 归一化的二维直方图在二维离散空间中的特征向量

系有：

- $D(h_i, h_j) = 0$ ，相等关系。
- $D(h_i, h_j) > 0$ ，一般关系。
- $D(h_i, h_j) = D(h_j, h_i)$ ，且 $h_i = h_j$ ，对称关系。
- $D(h_i, h_k) \leq D(h_i, h_j) + D(h_j, h_k)$ ，不等式关系。

§ 6.8 度量方法

本节中我们将给出基于直方图的颜色特征度量方法，主要的思路是根据距离度量。我们定义了一些有意义的距离，但只有一部分适合于相似性的度量。

6.8.1 弱相似性距

6.8.1.1 Minkowski 距(D_l)

Minkowski 距是一种弱相似性的测量，其中 h_q 表示查询直方图； h_t 表示目标直方图。

$$d_{q,t}^r = \left(\sum_{m=0}^{M-1} |h_q[m] - h_t[m]|^r \right)^{1/r}$$

当已知某一区域，查询含有与该区域相近所有图象。如果查询图象大小 $q <$ 目标图象大小 t ，且直方图并未归一化，则 $|h_q| \leq |h_t|$ 。此时直方图距为：

$$d_{q,t} = 1 - \frac{\sum_{m=0}^{M-1} \min(h_q[m], h_t[m])}{|h_q|}, \quad |h| = \sum_{m=0}^{M-1} h[m]$$

但是由于

$$d_{q,t} \neq d_{t,q}$$

所以上式经过修改得到下式:

$$d'_{q,t} = 1 - \frac{\sum_{m=0}^{M-1} \min(h_q[m], h_t[m])}{\min(|h_q|, |h_t|)}$$

则当 $|h_q| = |h_t|$ 时,有 $D_1(q,t) = d_{q,t}$

$$D_1(q,t) = \sum_{m=0}^{M-1} |h_q[m] - h_t[m]|$$

这是一个特例,实际上通过直方图归一化,可以满足以上条件。

6.8.1.2 欧几里得距(D_2)

欧几里得距是 $r=2$ 时求得的 Minkowski-form 距,

$$D_2(q,t) = D_2^2 = (h_q - h_t)^T (h_q - h_t) = \sum_{m=0}^{M-1} (h_q[m] - h_t[m])^2$$

将上式分解 v_2 为:

$$D_2(q,t) = h_q^T h_q + h_t^T h_t - 2h_q^T h_t,$$

则当 $|h_q| = h_q^T h_q = 1$,且 $|h_t| = h_t^T h_t$,我们有

$$D_2(q,t) - 2 = -2h_q^T h_t$$

因为直方图进行 $r=2$ 的归一化后,能满足上面的条件,所以使用此方法,将问题简化到只求两个直方图的向量内积。

6.8.1.3 直方图的余弦距 (D_2')

与欧几里得最相近的是余弦距。余弦距通常用来度量文本的相似性(文献[21]),余弦根据两个向量的夹角的余弦值计算它们之间方向的差异,作为其距离度量。

由于两各项量之间的内积可以表示为:

$$h_q \cdot h_t = h_q^T h_t = |h_q| |h_t| \cos\theta$$

由此可得余弦距

$$1 - \cos\theta = 1 - \frac{h_q^T h_t}{|h_q| |h_t|}$$

同时,由于 $|h_q| = h_q^T h_q$, $|h_t| = h_t^T h_t$,我们有

$$D_2' = 1 - \cos\theta = 1 - \frac{|h_q| + |h_t| - D_2}{2|h_q| |h_t|}$$

利用欧几里得距,当 $|h_q| = |h_t| = 1$ 时,余弦距可以简化为:

$$D_2' = 1 - \cos\theta = \frac{D_2}{2}$$

用图 6-7 来比较欧几里得距与余弦距。图 6-7 (a) 中, $|h_0| = |h_1| = 1$; 图 6-7 (b) 中, $|h_0| > |h_2|$ 。从图 6-7 (a) 与图 6-7 (b) 看,余弦距与向量长度无关,只与向量方向有关系;而欧几里得距既与向量长度有关,又与向量方向有关系,计算 $D_2(h_0, h_1)$ 与 $D_2(h_0, h_2)$ 的关系较

复杂；而余弦距只与方向有关，即得 $D_2'(h_0, h_1) = D_2'(h_0, h_2)$ 。

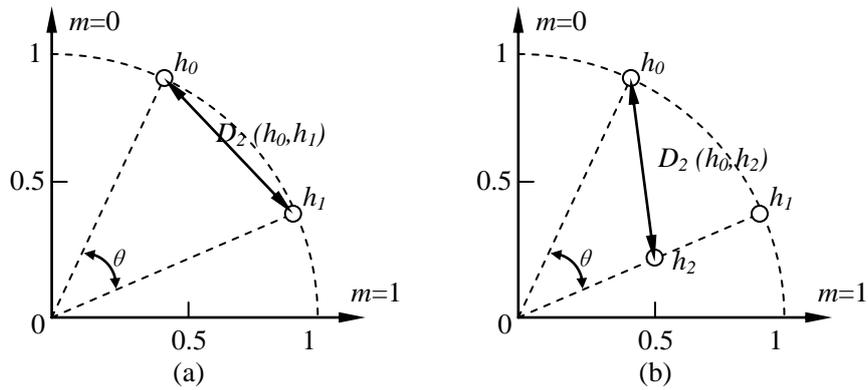


图 6-7 欧几里得距 vs 余弦距

6.8.1.4 二值集的海明距(D_3)

这里我们考虑用二值集近似表示直方图的情况。前面我们曾给出了用二值集逼近直方图的映射算法。设 s 为 M 维二值空间中的向量， $s \in B^M$ ， s 是一个二值向量，根据快速二值集映射算法我们得到，

$$|s| = \sum_{m=0}^{M-1} s[m]$$

由余弦距公式 $1 - \cos\theta = \frac{D_2}{2}$ ， $D_2 = 2 * (1 - \frac{s_q s_t}{|s_q| |s_t|})$ 得到，

$$D_3(q, t) = \frac{|s_q - s_t|}{|s_q| |s_t|}$$

由于 s 是一个二值向量，海明距可以通过向量不同位来确定，即通过对向量求异或得到：(如图 6-8 所示)

$$D_3(q, t) |s_q| |s_t| = s_q \oplus s_t$$

海明距的计算速度快，广泛应用在大型图象数据库中。

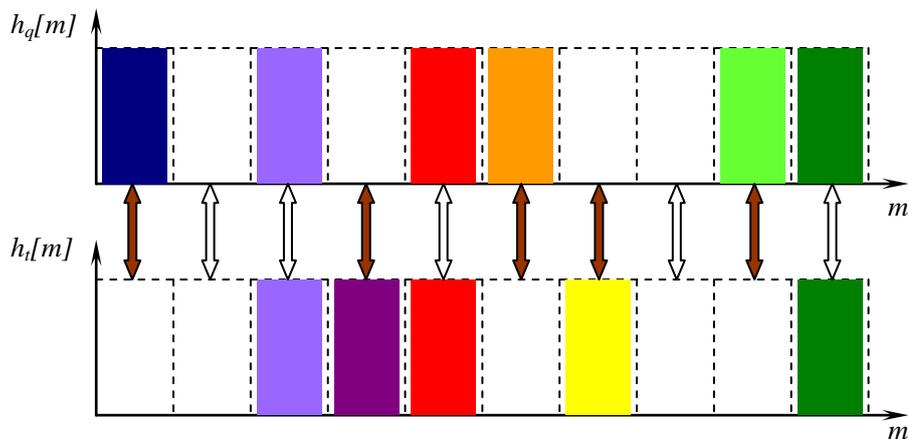


图 6-8 海明距可以通过向量求异或快速求得

上面这些度量方法之所以称之为相似性距，是因为弱相似性的测量存在问题，问题的关键在于所有这些方法都是通过直方图的分量逐一比较产生的（如图 6-9）。举一个简单例子：一个墨红相间的物体，究竟与一个红色物体更接近，还是与一个兰色物体更接近？答案为前者，这是显然的。但用弱相似性的办法，得到的结果可能是与两者接近程度一样，因此，需要对以上方法进行改进。

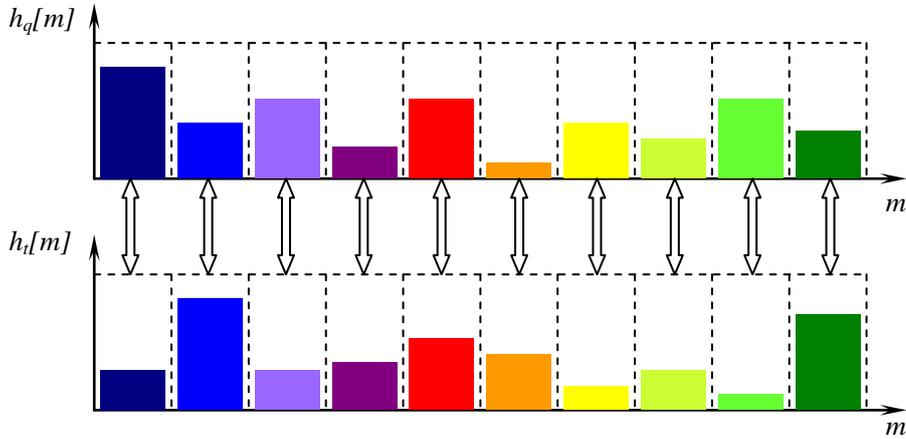


图 6-9 对直方图的分量逐一比较产生 D_1, D_2, D_2', D_3

6.8.2 相似性距

6.8.2.1 二次距 (D_4) 计算:

前面的相似性算法只是直方图的一对一的特征比较，实际一幅图象的某一颜色特征与相邻颜色都有关系，二次距 (D_4) 将采用一对多的算法来比较不同直方图。(如图 6-10 所示)

$$D_4(q,t) = D_4^2 = (h_q - h_t)^T A (h_q - h_t)$$

我们设 $A=[A_{ij}]$, A_{ij} 表示 h_q 中第 i 个元素 $h_q[i]$, 与 h_t 中第 j 个元素的相似程度。 $A_{ij}=A_{ji}$, 且 $A_{ii}=1$ (参见公式 12)。

在 IBM 的 QBIC™系统中便采用了二次距的方法度量直方图的相似性（文献[15]），此算法在比较效果上明显优于前几种，但在实际应用中，计算 A 矩阵需要大量的时间，因此还需要对该算法根据应用环境做进一步改善。

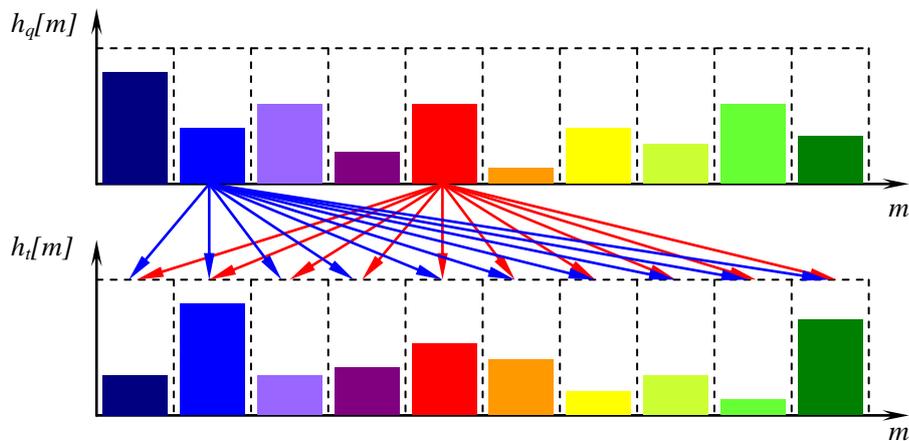


图 6-10 D_4 采用一对多的算法来比较不同直方图

6.8.2.2 二值集的二次距 (D_5) 计算:

二次距同样可以用来度量二值集的相似度。二值集的二次距可以由下式计算:

$$D_5(q, t) = D_5^2 = (s_q - s_t)^T A (s_q - s_t)$$

我们定义 $\mu_q = s_q^T A s_q$, $\mu_t = s_t^T A s_t$, $r_t = A s_t$, 则

$$D_5(q, t) = \mu_q + \mu_t - 2s_q^T r_t$$

此方法的优点是缩短了计算时间, 又提高了查询质量, 因此本文所涉及的工作采用了该方法来进行较精细的颜色相似性度量, 而较粗略的相似性度量则采用更为快速的海明距 D_3 。正如在第三章中所介绍的那样, 我们将 D_3 和 D_5 的运算通过数据库服务的存储过程实现, 以提高服务效率。

6.8.2.3 Mahalanobis 距 (D_6)

Mahalanobis 距是二次距的一个特例, 其中的变换矩阵 A 由一系列的训练用的直方图得到的协方差矩阵给出:

$$A = \Sigma^{-1}$$

我们将直方图向量看作一组随机变量:

$$X = (x_0, x_1, \dots, x_{M-1})$$

设 $R = [r_{ij}]$, $r_{ij} = E\{x_i x_j\}$, 这里 $E\{y\}$ 表示随机变量的 y 均值。那么协方差矩阵 Σ 可由下是给出:

$$\Sigma = [\sigma_{ij}^2], \text{ 其中 } \sigma_{ij}^2 = r_{ij} - E\{x_i\}E\{x_j\}$$

设 $x_q = h_q$, $x_t = h_t$, 则直方图与之间的 Mahalanobis 距可以表示为:

$$D_6(q, t) = (x_q - x_t)^T \Sigma^{-1} (x_q - x_t)$$

在特定的条件下, 如果 x_i 具有统计独立性, 但具有不均衡的方差, 则 Σ 将是一个对角阵 (文献[163]):

$$\Sigma = \begin{bmatrix} \sigma_0^2 & & & 0 \\ & \sigma_1^2 & & \\ & & \ddots & \\ 0 & & & \sigma_{M-1}^2 \end{bmatrix}$$

这种情况下, Mahalanobis 距可以简化为:

$$D_6(q, t) = \sum_{m=0}^{M-1} \left(\frac{x_q[m] - x_t[m]}{\sigma_m} \right)^2$$

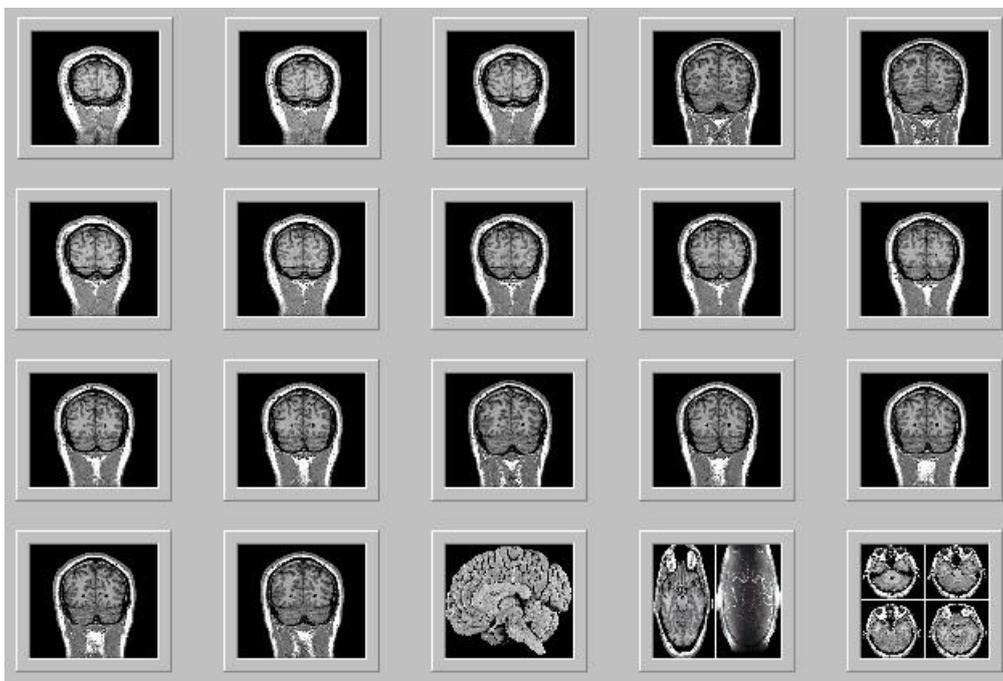
如果 x_i 不具有统计独立性, 可以通过坐标变换转换至一个具有统计独立性的坐标系中, 这可以将其看作一般的特征值问题加以解决。通过求得特征向量 φ , 使得 $\lambda\varphi = \Sigma\varphi$, 这样我们有 $\varphi Y = X$, 从而 Mahalanobis 距可以表示为:

$$D_6(q, t) = \sum_{m=0}^{M-1} \left(\frac{y_q[m] - y_t[m]}{\lambda_m} \right)^2$$

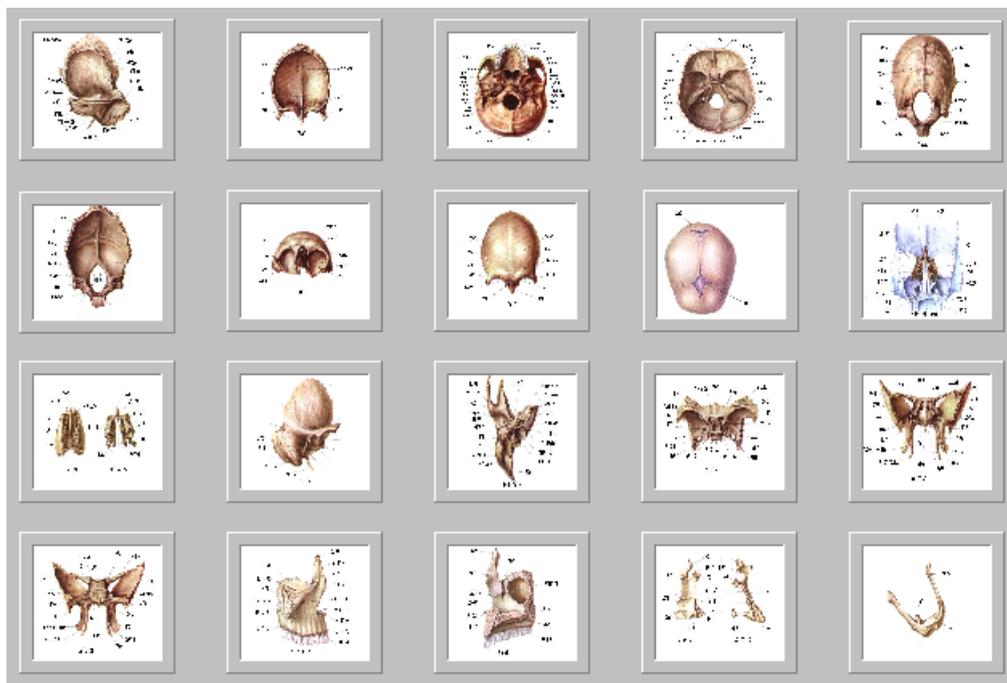
由于协方差矩阵可以通过不断的训练和学习得到, 所以可以用让用户培养检索过程以使得检索更加符合用户的相似性观点和习惯, 使得检索过程智能化。Mahalanobis 距是一种很有前途的相似性度量手段。但由于其计算上的复杂性, 在规模的图像数据库的检索上存在着效率上的问题, 所以不能用于大范围的快速搜索, 需要先使用其它技术缩小搜索范围, 然后再进行较精确的匹配。

§ 6.9 实验效果

我们对包含有 7000 余幅图像的数据库做实验，其中大部分为医学影像（90%），为了检验本文的方法对一般图像的有效程度，还有一部分普通的风光或人物画片和照片。在返回的前 20 幅图像中，第一幅为示例图像，其余的图像按与其相似的程度排序。检索效果如图 6-11 所示，图中相似顺序为先左右后上下。



(a) 灰度 CT 切片检索效果



(b) 彩色教学图像检索效果

图 6-11 图像示例检索效果（7000 余幅图像的数据库，医学影像占 90%左右）

由于采用了有效的特征表示和相似度比较算法,检索的效率和效果都是令人满意的,客户端一般能在小于 1 秒的时间内得到返回结果。我们的实验环境为:

CPU: PentiumII 266;

内存: 64M;

操作系统: Windows98;

数据库服务: Sybase Anywhere 5.0;

网络环境: 10M 带宽的 TCP/IP 局域网。

§ 6.10 小结

我们在本章中给出并评价了多个颜色空间的变换 T_c 和量化 Q_c 方法,主要的目标是获得能够有效定义颜色直方图的颜色特征空间。我们给出了 HSV 颜色变换,以及将其量化成 194 个颜色值的详细过程和方法,因为通过比较多个变换和量化方法,只有这种颜色空间是一致、完整、紧凑和自然的。

为了描述图像或区域中 HSV 颜色的分布,我们定义了颜色直方图 (h_c),同时给出了它的二值颜色集表示 s_c 。二值颜色集 s_c 仅描述了诸多颜色的一种筛选,以对图像和(或)区域颜色提供紧凑的度量方式。我们还给出了一个快速算法(BSFM)用以计算与给定颜色直方图最相近的二值颜色集。

随后我们给出了多种基于直方图的相似性度量方法进行了评价。我们最终选用了基于二值颜色集的海明距和二次距分别用于满足不同粒度的相似性检索需求。

除了本章叙述的工作外,我们进行了一些相关的研究。类似于颜色特征表示,我们还针对纹理特征和形状特征研究了直方图和二值集,这使得我们能够对这几种重要的视觉因素使用一种通用的、兼容的特征抽取、索引和检索方法。

第七章 基于 Web 的图像检索

医学影像往往具有分布式存放的特性,应充分考虑远程教学与诊断、甚至计算机协同工作的需要,提供局部和远程尤其是 Internet 访问支持,可以考虑 Client/Server、Web/Server 等数据库访问方式。尤其需要注意的是如何建立高效的医学影像的数据远程访问服务,将是本题目不同于一般 Internet 数据库服务的重要研究课题。

在第二章中,我们曾分析了各种数据库服务的体系结构,并选用了三层 Client/Server 结构,其中一个很重要的原因是因为这种结构适于在 Web 上实现。需要指出的是,我们在客户端可以完全不用浏览器而是用操作系统提供的 GUI 实现表示层的逻辑,此时在表示层/功能层/服务层之间通过 TCP/IP 协议进行通讯,功能层可以与表示层或服务层同处于一个物理节点,或三者相互独立。这样系统可以不加修改的运行于 Internet/Intranet。

然而,Internet/Intranet 之所以得到广泛的应用和发展的主要原因不只是由于它支持 TCP/IP 协议,更重要的是,在此协议基础上还支持 HTTP、FTP、Telnet、POP/SMTP、Gopher 等多种协议和服务,尤其是以 HTTP(超文本传输协议)、HTML(超文本标注语言)、MIME(多用途 Internet 邮件扩展协议)和 URL(统一资源定位)等协议为基本协议组的 Web 服务为访问 Internet/Intranet 上的信息提供了十分方便的手段。

为医学图像数据库建立基于 Web 服务可以大大提高数据库的可用性并拓宽信息的服务范围。客户端采用浏览器实现表示层逻辑,使得客户端可以独立于操作系统,操作简单直观,信息表现形式生动。以此为基础可以实现远程诊疗与教学,有很高的社会和经济效益。

实现 Web 数据库应用一般采用两种做法,一种是把应用程序下载到客户端并在客户端直接访问数据库,另一种是在 Web 服务器端提供中间件来连接 Web 服务器和数据库服务器。本章将首先简要分析这些技术。

§ 7.1 客户端直接访问数据库

为了完成数据库访问,可以通过 Web 浏览器把应用下载到客户端运行,在客户端直接访问数据库,客户端应用包括 Java Applet、ActiveX、Plug-in 等。一些 RAD(Rapid Application Development)工具支持这些可下载组件的开发,并且为了提高运行效率,这些 RAD 工具往往采用多种不同的技术扩充客户端和服务端的功能,形成不同形式的多层次服务结构。

JAVA 语言是 SUN 公司推出的第一个专门用于 Internet 和 WWW 环境的、面向对象的编程语言,是一种解释性语言。JAVA 语言的推出给 Web 开发提出了一个崭新的概念,原来是依靠服务器和 CGI 程序处理所有的任务,现在可以利用 JAVA Applet 在浏览器上分担相应的任务。JDBC 是 JAVA 语言的 SQL API, JDBC 支持 JAVA 程序和数据库系统之间实现无缝连接。JDBC 为 JAVA 编程者提供了与 ODBC 相似的功能,它是一个与数据库系统独立的 API,包含两部分: JDBC API 和 JDBC Driver API。使用 JDBC,可以将 SQL 语句传给几乎任何一种数据库,用 JAVA 编写的程序能自动将 SQL 语句传给相应的数据库服务器。JAVA 语言的与平台无关性使得它与 JDBC 结合所开发的应用系统能够真正实现“Write Once, Run Everywhere!”(文献[32])。

我们没有采用这种客户端直接访问数据库的方式来实现医学影像数据库的 Web 服务,因为这种结构从本质上来说只是把浏览器作为实现客户端功能的二层 Client/Server 结构。与我们在前面分析得到的三层结构不相符合,不利于功能与表示的分解。

但由于 JavaApplet 和 ActiveX 得到了大多数浏览器的支持,可以用来增强客户端的交互手段。医学影像数据库系统需要与图像有关的很强的交互,仅依靠浏览器提供的交互手段是不够的。我们在实际应用中,可以用 JavaApplet 或 ActiveX 增强表示层的交互能力,并借助于其他技术手段实现三层 Web/Server 结构。

§ 7.2 基于中间件的 Web 数据库服务

在第二章中我们曾介绍了三层 Client/Server 结构。基于中间件的 Web 数据库服务系统从本质上来看,是前面所述三层模式 Client/Server 系统的一种基于 Internet/Intranet 上 HTTP 协议的网络实现。它不仅具备普通两层模式系统的全部优点,而且还具有成本低和协议标准化的新特点,这使得 Web 数据库服务系统具有更好的性能/价格比。

基于中间件的 Web 数据库服务作为一种基于 HTTP 协议标准的多层 Client/Server 数据库服务模型模式,其基本框架如图 7-1 所示。

中间件负责管理 Web 服务器和数据库服务器之间的通信并提供应用程序服务(功能层),它能够直接或间接调用外部程序或脚本代码来访问数据库,因此可以提供与数据库相关的动态 HTML 页面,或执行用户查询并将查询结果格式化 HTML 页面,通过 Web 服务器返回给用户浏览器。

根据中间件的功能和结构以及数据库连接技术的不同,目前存在着多种 Web 数据库服务模型。最基本的中间件技术有通用网关接口(CGI)、应用程序编程接口(API)接口以及 Web_DB 产品。

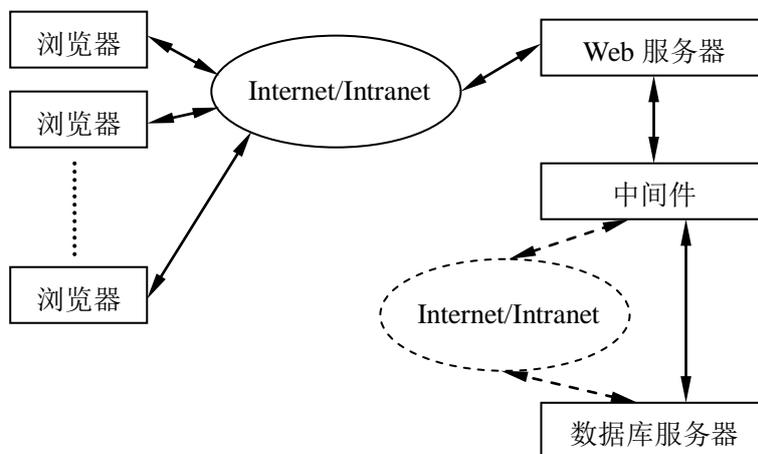


图 7-1 Web 数据库服务模型

7.2.1 CGI 技术

公共网关接口 CGI,及其改进产品 FastCGI,可以作为图 7-1 中的中间件,允许 Web 服务器执行外部程序,通过外部程序来访问数据库等外部资源,以产生 HTML 文档并返回 Web 服务器。这一方案的优点是易于实现,用户可以明用浏览器,暗用数据库。用户在 Web 页面上交互式填查询表,即可通过 Web 进行异地数据库存取,并在浏览器上显示结果。缺点是当中间件为 CGI 时,每次访问都需要重新启动一个 CGI 程序实例,它以进程形式运行,占用资源较多,效率较低,且难于解决连续事务处理。

7.2.2 API 技术

为了克服 CGI 局限, 一些 Web 服务器厂商开发出专用的 API, 允许开发人员编写程序来扩充服务器的功能, 如 Microsoft 的 ISAPI、Netscape 的 NSAPI 等。与 CGI 相比, API 应用程序与 Web 服务器结合更紧密, 占用的系统资源也少得多, 而其运行效率却大大得到提高, 同时还能提供更好的保护和安全性。

然而, 如果纯粹依赖于 API 实现数据库访问, 开发 API 应用程序也要比开发 CGI 程序要复杂的多, 而且, 由于各种 API 之间兼容性很差, 缺乏一个统一的业界标准, 因此应用程序往往还只能工作在专用的 Web 服务器甚至操作系统上。

7.2.3 Web_DB 技术产品

如上所述, CGI 和 API 有一系列缺点, 这些缺点造就了 Web 数据库的商业机会, 各公司竞相出台 Web_DB 技术产品。根据实现技术不同, 又可以分为以下几种:

(1) 提供数据库产品的 CGI 接口。

这类技术通常为数据库产品及其开发工具提供专用的 CGI 程序接口, 将开发人员从繁重的 CGI 编码中解放出来, 从而提高开发效率。如用于 Microsoft Visual FoxPro 的 FoxWeb。这类技术并未克服 CGI 的局限。

(2) 提供支持数据库访问的 API 接口。

这类产品在 Web 服务器的 API 基础上, 提供数据库的访问接口, 在客户端采用 VBScript、JavaScript、JavaApplet、Plug-in 等技术实现动态 Web 应用。目前这类产品最丰富, 应用也比较成熟。如 Microsoft 的 IDC、Borland 的 Web Broker、Sybase 的 Web.SQL 和 Web.PB 等。

(3) 提供专用的 Web 服务器, 捆绑数据库服务。

如 Oracle 的 WebServer 能和 Oracle7/8 数据库系统无缝结合, 使应用程序具有较高的运行效率。但难以连接其他数据库系统。

这些技术和产品各有特色, 其中很多中间件产品在充分兼容 CGI 标准和利用 API 技术的基础上, 还引入了扩充的 HTML 标集合服务器端的程序或脚本, 并允许建立包含扩充标记和脚本的 HTML 页面模板; 而在响应用户的请求的同时, 则通过转换这些特殊的标记和执行相应的程序或脚本来完成数据库的访问, 同时生成动态的 HTML 页面返回给用户。

但我们也注意到, 大多数的 Web_DB 产品是由 Web 服务器或数据库厂商开发的, 很少能够做到同时兼容不同的操作系统、Web 服务器和数据库系统, 并且又具有较好的开发和运行效率。

7.2.4 Cold Fusion

我们在实现基于 Web 的医学影像数据库检索服务时采用的 Allaire 公司的 ColdFusion 也属于 Web_DB 产品, 但是由于 Allaire 属于第三方开发商, 使得 ColdFusion 能够很好地独立于操作系统、Web 服务器和数据库系统。

整套 Cold Fusion3.11 包含三个部分:

Application Server。能在任何主要的 Windows 和 Solaris Web 服务器上工作, 发布具有高性能的 Web 应用程序。

Cold Fusion Studio。Cold Fusion 的开发工具。它使用图表式的数据库和编程工具快速, 简单地把浏览器, 服务器, 数据库技术集成到复杂的应用程序中。

Cold Fusion Markup Language (CFML)。该语言使用基于标志符的语法, 和传统的 HTML 非常的相似, 对于编制网页的新手而言, CFML 非常的易学。

Cold Fusion Application Server 为 Cold Fusion 中最为重要的一部分。它提供了所有开发者需要用于开发高级 Web 应用程序的强大的功能。Application Server 可以满足绝大部分网

站的需要。Cold Fusion Application Server 具有以下的特点:

- 多线程的服务。
- 对 SAP (Server Application Programming) 的支持, 其中包含 ISAPI, NSAPI, WSAPI 和 Apache API。
- 数据库连接池, 该功能可以提高数据库事务处理的性能。可以连接几乎所有的比较流行的数据库系统。
- 预编译页和缓存技术, 该技术可以消除磁盘 I/O 过程, 从而提高应用程序的性能。
- 和 Email Server, Web Server, Directory Server 和 File Server 的全面连接。
- 内置的搜索引擎, 为 Verity SEARCH'97 全文索引和搜索引擎。使用它可以实现对网站的内页检索。

由于我们将医学影像数据库的检索任务的核心工作放到了功能层面上, 这就需要在服务端的服务上完成大部分的工作。我们需要在一般的数据库中间服务上增加有关图像分析与处理等复杂的应用服务。Cold Fusion 的下述特性使得我们获得了实现这一目标的有效手段:

(1) ActiveX Server 元素。扩展的 Cold Fusion 能利用任何 ActiveX Server 元素 或 COM 或 DCOM 对象连接到任何外部目标, 调用它的任何方法以及在仅以类似于高级 CFML 语法的范围内编写程序来获得或设置其性质。可以为动态主页利用任何可提供的 Server 元素, 包括它们的 动态数据对象 (ADO), 可以完全适合于 Normandy 服务, Windows NT BackOffice 服务器, 以及第三方的控制应用, 可以很容易的利用 Visual Basic、Delphi、C++Builder、VC++、Java 等编写的过程元素来扩展 Cold Fusion。

(2) 开放 API。利用 Cold Fusion 应用程序接口 (CFAPI) 可以创建新的自定义 CFML 目标 (CFXs) 并能调用旧有程序、OLE 自动控制服务器、或诸如在线信用卡认证等功能实现的第三方应用。CFAPI 使用了高性能的 C++接口。

(3) 自定义目标。建立可重复调用的元素或‘自定义目标’那么就能压缩由 CFML、HTML 或其他客户端技术创建的复杂的过程和函数。在动态主页中利用数据库记录设置或其他的输入输出数据库设置的的变量。

§ 7.3 Web/Server 模式的主要问题与对策

我们在第三章中已经分析了图像检索在远程数据访问中存在的一些问题和相应的对策, 除此之外, Web/Server 模式的数据库服务模式目前存在的主要问题是:

(1) 网络带宽。因为用户端没有或很少有程序代码, 更没有最新的数据, 每次查询以及反馈度需要通过网络来和服务器取得联系, 势必占有一定量的网络通讯带宽。通过服务器及浏览器的页面缓冲技术以及本地文件技术等, 保证只传送必要的信息, 从而提高通讯速度, 降低通信成本。

(2) 效率及可靠性。HTTP 协议效率较低, 对于交互式处理的检索系统, 当通讯过程中发生瞬间线路断开时, HTTP 不具备重新通讯的能力, 而且每传送一个文件都要启动、结束会话。对于交互式处理, 为了提高启动/结束会话的效率, 就必须在服务器上设置大容量的缓冲。

(3) 安全问题。有效解决网络及信息安全问题的方法是利用“防火墙”技术, 将关键的核心信息资源与网络分离开来。

§ 7.4 基于 Web 的图像检索

在 Web 上提供图像检索服务已经是目前各 Internet 搜索引擎服务厂商提高竞争力、吸引网络访问的一种有力手段。我们首先考察一下这些流行的图像搜索引擎, 然后讨论这些图像搜索引擎的实现技术、在医学影像 Web 检索中值得借鉴和需要改进的地方

7.4.1 几种流行的图像搜索引擎

图像搜索引擎是专门用来在 Internet 上搜寻图像的搜索引擎, 同文字搜索引擎一样, 它提供一个分类目录, 并提供关键词检索功能。我们可以通过分类目录的浏览, 或者使用关键词检索, 来查找我们关心的图片。

与文字搜索引擎不同的是, 在这里查到的最终结果, 不是文字形式的网站名、网站内容提要或网页内容, 而是一幅幅微缩图片及其地址链接。

(1) Infoseek 的 Image Surfer。其分类目录共有 9 个大类, 72 个二级类目, 并由三级类目, 收录的图片非常丰富。(www.infoseek.com)

(2) Yahoo 英文版的 Image Surfer, 它是 Yahoo 的一个部分, 类目较前者少, 收录的图像也比前者少一些。(www.yahoo.com)

以上两个图像引擎, 使用的都是 Excalibu 出品的一个叫做 Image Surfer 的图像搜索软件。当我们查询分类目录, 浏览到最下层级别时, 会得到这样的结果:



图 7-2 Image Sufer 分类检索结果

在分类目录搜索结果页面上提供了一般的翻页浏览控制, 每页提供 6 幅图片。单击图片, 就会转到相应的网页上看到它原来的面目。“Image Info”链接有当前图片的更详细的超文本说明。

而真正体现图像搜索引擎特点的是“Visual Search”功能, 单击它, 用户可以得到 5 页 (每页 6 幅) 与当前图像看起来内容、形象相近的图片。

通过关键词搜索的结果与图 7-2 不同, 比如我们用关键词 dog 进行检索, 得到的结果如图 7-3 所示。

除了分页浏览方式有所改变以外, 值得注意的是, 此时不再提供“Image Info”和“Visual Search”功能。

类似的图形搜索引擎还有 Lycos 的 Pictures & Sounds, 哥伦比亚大学的 WebSeek 等。AltaVista 的图片检索是通过关键词中加限定词“image”实现的, 如“image:dinosaur”意为查找有关恐龙的图片。另外 Internet 上还有一些专业的图像数据库, 一般都采用分类目录的方法提供图像搜索服务, 像知名的 NIX (美国航空航天总署图像交流中心)、Smithsonian 图片数据库等。



图 7-3 ImageSurfer 关键词检索结果

7.4.2 Image Surfer 的实现技术

以 Image Surfer 为代表的图像搜索引擎在实现上主要综合采用了我们在第二章中所述的基于描述的直接查询（关键词搜索）、链接查询（分类目录和超链接）和直观浏览（分页浏览控制），是我们所给出的图像检索模型的简化实现（比较于图 2-2，如图 7-4 所示）。

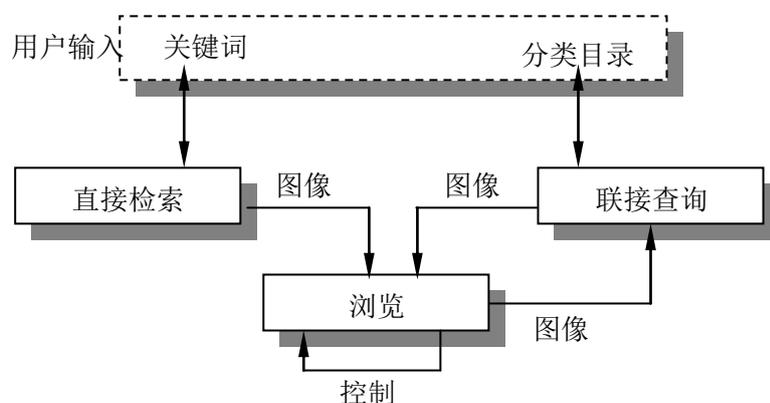


图 7-4 Image Sufer 图像检索流程图

无论是 Yahoo，还是 Infoseek，在其 Image Sufer 中提供了能充分体现图像检索特点的“Visual Search”功能，从检索行为和效果来看，似乎是一种内部图例查询的方式。但我们没有把它作为内部图例查询在图 7-4 中反映出来，是因为我们通过分析其查询过程和方式，认为只是一种链接查询方式。Image Sufer 如果能够支持在服务端根据用户选中的图例进行动态的图像分析和特征抽取，那么将很容易实现关键词检索后根据用户随意选定的图像实施相似检索，可惜的是我们没有看到这项特征。而且，通过分类目录查询的结果进行的“Visual Search”只能检索相似的 30 幅图片，这个数目是固定的，显然不能动态地根据与当前图像的相似度进行索引，由用户指定需要的最相似图像的数目。可见，与当前图像最相似的 30 幅图像是预先定义的，在数据库中通过关系，或通过直接超链接实现。

相似性预定义过程可以在图片插入数据库时进行，或由图像检索服务器中的一个相似性定义进程周期性地地进行，后者总能反映最近一段时间间隔内数据库中图像的相似状态，而前者一旦定义便不再改变，可以节约大量的时间和空间资源，但一段时间后便不能有效反映数据库的图像相似关系。较好的办法应该是综合利用这两种方法，首先在插入图像时定义相似关系，在一定的时间间隔后，由一个相似性定义进程更新那些最长时间未获更新定义的图片的相似关系，从而在效率和效果之间获得折衷。

7.4.3 值得医学影像 Web 检索借鉴的技术

虽然 Image Sufer 与我们在第二章中提出的图像检索模型还有一定的差距，但仍不失为

一种成功的、在实践中得到检验的技术。其中有一些我们在实现医学图像数据库的 Web 检索时值得借鉴的地方:

(1) Web 是一种以超文本、超链接为主要交互手段的一种技术,实践证明这种手段直观易学、简单有效。将这种技术应用于搜索引擎的检索服务时,通常表现为分类目录的方式。对医学图像,按照医学领域对医学影像的多种组织管理方式建立多种多层次的分类目录,将是用户最为熟悉、最可接受的方式,如可以按照管理影像的行政机构设置、影像涉及的医学学科、影像诊断出的病变类型等建立多层的目录结构,实现联接查询,以满足不同检索者的不同需要。

(2) 基于描述(关键字)的直接查询也是一种强有力的图像检索手段。尤其对于医学影像,由于影像的产生多以诊断、治疗、教学为目的,一般具有与之相联系的详尽的描述信息,如病人的病历、医生的诊断、病人手术记录、病人随访记录、教学的注解等,而且这种查询也反映了建立医学影像数据库最一般、也是最主要的目的和任务。基于描述的直接查询应成为 Web 医学影像检索服务的主要手段。

(3) 预定义的手段可以极大地加速基于内容的检索过程,降低检索的资源要求。一般说来,正如本文在前几章所述,图像的颜色、纹理和形状特征可以用一种基于直方图和二值集的方式进行定义,在用户进行的图例检索时,只对图例进行特征抽取,然后根据图例的特征和数据库中个图像预定义的特征的相似度排序,返回一定数目的相似图像。由于数据库中图像的特征已经预定义,所以在一般的相似性检索需求下,无需对数据库中的图像多次进行图像分析和特征抽取。

(4) 不仅特征需要预定义,当医学影像数据库达到一定的规模,或者数据库服务的同时访问客户达到一定的程度时,有必要进行图像之间的相似度也需要预定义。在 Web/Server 模式下,Web 服务器和数据库服务器的效率十分重要。当检索相似图像时,需要对数据库中全部或部分图像求取相似度并进行排序,如果数据库的规模很大,或者同时检索的客户过多,Web 服务器和数据库服务器将不堪重荷。用户对相似图像在数目上的要求一般不会太多(如 20 幅较相似的图像),那么我们可以对每一幅图像在当前数据库中求取最相似的 20 幅图像并为之建立联接,用户在检索时直接返回这 20 幅图像。这样可以用一定的空间代价换取极大的效率

(5) 由于基于内容的检索存在着效率上的问题,加上医学上的图像应用一般是明确的而不是模糊的,所以应以联接查询和基于描述的直接查询为主要检索手段,缩小检索范围后再辅之以基于内容特征的检索。

§ 7.5 小结

本章中我们简要介绍了几种基于 Web 数据库应用技术,经过分析与对比,我们认为适于医学影像数据库 Web 检索服务的模式是基于中间件的三层 Web/Server 模式的数据库服务,并介绍了我们在实践中应用的 Cold Fusion。

Web/Server 模式存在着网络带宽、安全性、效率和可靠性等方面的问题,对此我们简要地给出了一些对策。

通过对目前流行的一些图像搜索引擎的实现技术的分析,我们在最后给出了医学影像数据库的 Web 检索服务的实现中值得借鉴的一些技术。

第八章 总结及展望

本文主要就近两年来我们在医学影像数据库方面一些工作和研究以及其中的一些技术要点如图像检索模型、系统结构与数据库结构、图像处理与分析、基于内容的检索和基于Web的检索等作了一些介绍和回顾。可以简单地用一句话来概括医学影像数据库的研究内容,那就是充分考虑医学领域需求的图像数据库研究。但还不仅如此,医学影像数据库是建立在图像数据库、图像处理与理解、人工智能、远程与局部网络技术以及医学领域知识基础上,支持医学影像数据有效存储、传输、检索和管理的数据库技术。

医学影像数据库研究中一个基础的、重要的、有意义的工作是图像数据库的检索技术研究。目前图像数据库方面的研究无论在国内还是在外国都处于一种方兴未艾的局面,我们在这方面的时间也不是很长。我们认为图像数据库技术的研究一方面要尽量与领域无关以寻求广泛适用的技术和方案,但另一方面,理论的研究不能脱离实践的需要和检验,所以医学对影像管理的迫切需要使得我们获得了开展这项研究的土壤和目标,从而可以对一些通用的技术和方案进行验证和实施,并针对领域的特殊需要,对我们的这些技术进行提炼和改造。提炼的目的并不是将领域的需要和通用的模型进行揉合,而是使之更加分离,使得通用的技术跟加抽象和普遍,从而为在其它领域进行图像数据库技术的应用建立可靠的、经过检验的基础。

就本文所涉及的图像数据库技术而言,还有很多有待发展和深入的地方。首先,本文提出的基于直方图和二值集的内容特征的提取和抽象方法,虽然是建立在对颜色、纹理和形状三种特征的全方位的研究与考察上,试图建立一种对不同特征的统一的观点和解决方案,但是到目前为止,我们仅对颜色特征进行了详细的落实和实现。要将这种方法进一步落实到纹理和形状,其中必然还有一些具体的细节问题需要作进一步的研究和探讨。

从本质上来说,直方图和二值集实际上是对特征的统计表达和压缩。当然在基于内容的特征表达方面还有一些不同的方法,如对纹理的基于结构的方法(文献[1])、对形状位置关系的拓扑表达方法(文献[19])等,如何在图像数据库中提高这些方法对用户的透明性和可用性仍是值得研究的课题。

我们的医学领域的图像数据库是实现在完备的关系数据库基础上的,对数据库系统本身的要求除了具有较完善的关系能力和图像数据的存取能力外没有其它特别的要求,虽然可以通过中间件完成特征抽象和内容检索,可以在现有的技术条件下获得较高的灵活性和独立性,但是必须为此付出中间件和数据库服务之间的通讯代价。根本的出路是数据库本身支持可扩充的多媒体数据的表示和检索,这就需要数据库从底层支持用户自定义的数据类型和数据操作。目前的对象-关系数据库研究便是针对这样的需求进行的。

图像是一种格式化的数据(Data),但不包含有关图像内容的格式化的信息(Information),如果图像本身能够对其内容进行自阐述,将使得图像检索的工作相对简单。目前正在制订的MPEG4和MPEG7中针对视频检索和交互点播的要求加入了有关内容的信息,这对静态图像也是值得借鉴的。但从本质上来说这是一种特征预定义的方法,有一定的局限性。

回顾一下我们在绪论中关于医学影像数据库中需要研究的内容和要解决的问题的叙述,并对比我们已经开展的工作,也同样有许多方面需要开展或进一步开展研究。如数据库与医学影像设备的接口问题、医学领域知识的导引作用、在基于内容的医学影像检索中医学内容特征的特殊作用等。

在过去的工作过程中,我们对基于病变的纹理和器官的形状进行内容检索探讨了其可

能性，这是一个很有吸引力和挑战性的题目，因为这时领域知识的导引作用将占主导地位，领域分析与领域知识表达和作用机制将是这个题目的主要研究课题。

把目光再从图像数据库上放开去，我们将看到更广泛的多媒体数据库，其中不仅包括图像、可能还包括视频、声音等其它信息表达媒体。基于内容的检索（CBR, Content Based Retrieval）将使得我们面对多种不同方面的特征以及这些特征之间的关系，多媒体数据将对传统数据库的数据模型从数据结构、数据操纵和数据完整性等各个方面提出新的要求，传统的数据模型，包括现在广泛应用的关系模型，在管理这些数据时都将勉为其难。在这方面存在着许多富有挑战性的研究方向需要我们将进行更加深入的工作。

如此多的挑战和困难并不能阻止我们对美好未来的憧憬：我们用孩子小时候的一幅照片或者一段“小老鼠，上灯台”的录音检索家庭录像资料库，计算机将自动为我们做好一个“幸福的童年”专题集锦。如果说这个应用距离我们还太遥远，那么，也许明年元旦我们就可以从 Internet 上的某个大容量图像库中很容易查找一幅满意的风景画片，比如“2000 年的第一个海上日出”，作为电子贺卡的画面或 Windows2000 的画布，来庆贺一个新世纪的到来。

参考文献

1. 王润生。图像理解。长沙：国防科技大学出版社，1995。
2. 赵荣椿等。数字图像处理导论。西安：西北工业大学出版社，1996。
3. 李向阳，鲁东明，潘云鹤。基于色彩的图像数据库检索方法研究。计算机研究与发展，1999，36（3）：359~363。
4. 曹莉华，柳伟，李国辉。基于多种主色调的图像检索算法研究与实现。计算机研究与发展，1999，36（1）：97~100。
5. 伯晓晨，刘健平。基于颜色直方图的图像检索。中国图像图形学报，1999，4（1）：33~37。
6. 邬浩，潘云鹤，庄越挺，杨宇挺。基于对象形状的图像查询技术。软件学报，1998，9(5)：343~349。
7. 萨师煊，王珊。数据库系统概论。北京：高等教育出版社，1990。
8. K.R. Castleman. Digital Image Processing. Prentice Hall, Inc. 1996.
9. J.d. Ullman, J. Widon, A First Course in Database Systems. Prentice Hall, Inc. 1997.
10. D. Oliver, S. Anderson, J. McCord, S. Gumas, B. Zigon. Tricks of The Graphics Gurus. Sams Publishing, 1993.
11. J.R. Smith and S.F. Chang. Automated binary texture feature sets for image retrieval. In Proc. Int. Conf. Acoust., Speech, Signal Processing, Atlanta, GA, May 1996. IEEE.
12. W.I. Grosky. Multimedia Information Systems. IEEE Multimedia, 1(1):12-24, 1994.
13. R. Jain. Infosopes: multimedia information systems. In Multimedia Systems and Techniques, B. Furht, ed., Kluwer Academic Publishers, Boston, 1996, pp. 217-253.
14. N.S. Chang and K.S. Fu. Query-by-pictorial example. IEEE Transactions on Software Engineering, SE-6(6):519-524, 1980.
15. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele and P. Yanker. Query by Image and Video Content: The QBIC System. IEEE Computer, 28(9): 23-32, 1995.
16. K. Srihari. Automatic Indexing and Content-Based Retrieval of Captioned Images. IEEE Computer, 28(9): 49-56, 1995.
17. W.C. Huibers. An Axiomatic Theory for Information Retrieval. PhD Thesis, Universiteit Utrecht, The Netherlands, 1996.
18. Marchionini and B. Shneiderman. Finding facts versus browsing knowledge in hypertext systems. IEEE Computer 19, 1988, pp. 70-80.
19. E.G.M. Petrakis, C. Faloutsos. Similarity Serching in Medical Image Databases. IEEE Transactins on Knowledge and Data Engineering 9(3):435-447, 1997.
20. J.R. Smith, S.F. Chang. Quad-Tree Segmentation for Texture-Based Image Query. Preceeding of the 2nd Annual ACM Multimedia Conference, San Fracisco, Ca. October, 1994.
21. I.H. Witten, A. Moffat, and T.C. Bell. Managing Gigabytes: compressing and indexing documents and images. Van Nostrand Reinhold, New York, NY, 1994.
22. M.A. Oliver, N.E. Wiseman. Operatioons on Quadtree Encodeed Images. The Computer Journal 26(1):83-91, 1983.
23. C.A. Shaffer, H. Samet. Optimal Quadtree Construction Algorithms. Computer Vision, Graphics, and Image Processing, 37:402-419, 1987.
24. G.M. Hunter, K. Steiglitz. Linear Transformation of Pictures Reperesented by QuadTrees.

- Computer Graphics and Image Processing, 10:289-296, 1979.
25. 周明全等, 三维医学可视化系统的设计与实现, 西北大学学报, 1999, No6, Vol29, p1-4。
 26. 耿国华, 周明全, 三维医学可视化平台的研究, 微机发展, 1999,, Vo45, p55-57。
 27. 周明全等, 可视化技术及其应用, 西北大学学报, 1996, No6, Vol26, p349-353。
 28. 周明全等, 计算机辅助的颅骨面貌复原技术, 西北大学学报, 1997, No5, Vol27。
 29. 耿国华, 周明全, 一种从空间物体到八叉树转化的简捷算法, 西北大学学报, 1996, No4, Vol 26。
 30. 冯筠, 多媒体信息系统基于内容检索技术的研究和设计, 西北大学学报, 1998, No1, Vol 28, P17-20。
 31. 刘硕等, 模糊数学在图像增强中的应用, 西北大学学报, 1999, No4 (待发表)。
 32. 孙朝晖, 耿国华, 周明全, 基于 WWW 的信息系统的连接技术及其一种实现, 西北大学学报, 1999, No3 (待发表)。

致 谢

衷心地感谢我的导师耿国华教授。

耿老师是我目前为止所遇到的事业心最强的老师，她所表现出的认真、严谨、踏实的治学作风和废寝忘食的工作精神为我所深深敬佩。在三年科研和工作中，她给了我很多有益的指导和帮助，不仅如此，还亲自着手一些十分细致工作，我有多次配合她调试和修改程序的经历，使我看到了什么叫对用户负责，什么叫物我两忘。耿老师曾教导我们：“做学问先做人。”她做到了。

医学影像数据库的研究工作是在耿老师的带领和组织下进行的，在我的论文选题上，耿老师给予了充分的支持。耿老师在数据库、数据结构和人工智能方面有着深厚的见地和修养，使得我在研究和论文工作中获得了许多宝贵的知识和建议。

衷心地感谢周明全教授给与我一贯的指导、支持、鼓励和关心。他旺盛的工作精力和不屈不挠的工作态度以及对生活的乐观态度一直令我佩服不已，也是值得我学习的。

特别感谢耿老师和周老师给予我的众多研究和学習机会，在他们的悉心指导和关怀下，我有机会参加了计算机辅助的法医人类学颅骨面貌复原（国家“九五”重点科技攻关项目）、三维医学可视化平台研究、三维地层地质模型系统、多媒体人脑解剖、交通信息综合管理系统、医学影像数据库研究以及与日本大学关于数据库安全方面的短期合作等工作，获得了广泛的锻炼和培养。

特别感谢耿老师和周老师对我及我的家人在生活和工作上的关心。感谢他们对我必须在单位、家庭和学校之间分配时间的容忍。

感谢所有的研究生任课老师对我的教诲和关心。华庆一老师的风趣幽默、卞雷老师的循循善诱给我留下了深刻的映像。我清楚地记得洪蓬老师强忍胃痛，靠在讲台上坚持讲课的情景；也不会忘记杨留记老师在得知我住得较远后，让我不要为了赶早上第一节他的课而忽视交通安全，骑自行车不要太快了。老师们所教予的知识将是我修身立命的根本，将使我终身受益匪浅。

感谢我们的研究生秘书安娜老师的辛勤工作以及对我在工作和生活上的关心。

在本文的论述中涉及有关的实际工作时，我总是使用“我们”一词。这说明几乎所有的工作都离不开一个群体的努力。除了耿老师和周老师外，有很多的同学参与了研究和讨论，在此一并致以感谢。特别感谢晏春莉同学，她参加了大部分的研究工作并对基于颜色的图像检索进行了程序实现。特别感谢樊宏斌同学，在我的要求下，他对CT数据和分割结果向数据库要求的格式的转换，在很短时间便完成了程序实现。

特别感谢我的家人，尤其是我的妻子徐雯。三年中她承担了大部分的家务和教育孩子的义务，给我在生活上以无微不至的照顾。没有她的理解和支持我是无法完成这三年的学业的。

最后还有一个重要的人物需要感谢——我的儿子胡水，虽然他只有四岁，但在我需要用计算机写我的这篇论文的时候，不管是什么时候，哪怕他正在用计算机播放他最喜欢的卡通片《Tom&Jerry》，总是非常慷慨地让出计算机。而且，从我开始写论文起，他已经很长时间没有要我给他用“画笔”画房子了。