

●蔡巍 刘玉照

基于内容检索多媒体数据库的构建

摘要 研究了图像、视频等媒体的基于内容索引方法及如何开发基于内容多媒体检索系统模块,嵌入到面向对象多媒体数据存储库中,从而完成基于内容检索多媒体数据库的构建,并提出该系统的总体结构。图1。参考文献29。

关键词 基于内容检索 多媒体数据库 数据库建设

分类号 G258.94

ABSTRACT In this paper, the authors discuss some aspects in the construction of content-based multimedia databases, such as indexing methods, module designing and system structure. 1 fig. 29 refs.

KEY WORDS Content-based search. Multimedia database. Database development.

CLASS NUMBER G258.94

1 基于内容检索与多媒体信息查询

多媒体信息查询问题随多媒体信息系统的应用广泛而日益突出起来,特别是随着近年来信息高速公路(因特网提供了大量分布式的多媒体信息内容)的兴起,对这一问题的解决要求更为迫切。

多媒体数据包含有十分丰富的信息内容,其中图像、视频、音频等众多媒体类型更有复杂的语义特征,以往那种在关系数据库框架内使用赋以关键词或原子属性而进行的检索^[1]是一种浅层次的信息查询,不能揭示出更有价值的信息,造成极大的信息资源浪费。

基于内容检索(Content-based Retrieval)^[2],就是从媒体数据中提取出特定的信息线索,然后根据这些线索从大量存储在数据库中的媒体中进行查找,检索出具有相似特征的媒体数据来。它可以在更深层次和更有效地利用存储的多媒体信息。

例如,一段波涛汹涌的画面,配以不同的背景音乐,可有多种寓意。单纯使用关键词法检索,就无法从多侧面、多角度、多层次查询有关信息,而使用人工浏览又浪费时间。但若提取出媒体中的信息线索(如图像的形状、轮廓、

颜色等),基于这些特征化线索,进行相似性匹配,则易取得较好效果。媒体信息的特征化,是对媒体信息内容较深语义层次的揭示。

基于内容检索是一个逐步求精的过程^[3]:

(1)初始化检索说明。用户检索时,最初可用基于示例的查询或查询语言来形成检索格式。系统提取示例的特征或把查询描述映射为具体的特征。

(2)相似性匹配。将查询特征与特征库中的资料按照一定的匹配算法进行相似性匹配。满足一定相似性的一组候选结果按相似度大小排列返回给用户。

(3)特征调整。对系统返回的查询结果,用户可选择满意结果,或从中选择一个示例,经特征调整,形成新的查询。

(4)逐步缩小检索范围,直到用户满意。

基于内容检索时根据媒体特征进行相似性匹配检索的媒体特征有^[4]:颜色(color)、纹理(texture)、轮廓(sketch)、形状(shape)、空间约束(spatial constraints)、动态(motion)、领域概念(domain concepts)等。

基于内容检索是一门信息检索技术。在开发基于内容检索系统时涉及多学科的知识,

包括数据库管理系统、自然语言理解、图像理解、视频处理、模式识别、认知科学、计算机图形学、计算机视觉、信息检索等。基于内容检索与图像理解、模式识别等的区别在于它不是去理解图像的内容或识别它的内容,而是利用这些相关学科的综合知识去实现多媒体信息的基于内容的快速查找。

目前,基于内容检索技术在多媒体信息系统的主要应用领域有:艺术画廊和博物馆管理、建筑与工程设计、地理资源遥感与管理、地理信息系统^[5]、商标及版权数据库管理^[6]、房地产信息系统及面部检索系统、数字视频图书馆^[7]、WWW 信息浏览^[8]、多媒体 CAI 应用^[9]、多媒体出版^[10]等。随着该技术的逐步成熟,应用前景会更加广阔。

2 基于内容的图像表示和检索技术

基于内容检索,先前的方法可分为两种^[11]:(1)图像内容以手工抽取的一套属性建模,在传统数据库管理系统框架内管理。图像的基于属性表示是高级图像抽取的结果,数据库研究者一般使用该方法。(2)依靠一种集成的特征抽取—对象识别子系统以克服基于属性检索的局限。当图像插入数据库时,该子系统自动进行特征抽取和对象识别。该方法是图像理解学者提出和发展的。但两种方法均有缺陷。第一种是抽象程度不能太高,越高则图像查询的范围越窄;第二种是计算费用高、困难大且受限于专有领域。现在的方法是融合这两种方法的优点,克服它们的缺陷。查询处理使用媒体特征,特征的抽取(需要专家系统技术^[12]的辅助而实现)是自动或半自动的。

图像存储与检索的大致流程如下:(1)存储部分,使用扫描仪等装置将图像等输入计算机,进行图像分析,提取有关媒体特征,并实现图像及特征的存储,形成图像数据库;(2)检索部分,将要查询的图像使用扫描仪输入计算机,进行图像分析并提取媒体特征,然后利用这些特征与图像数据库中媒体特征进行相似

性匹配,最后输出符合条件的图像。

图像媒体相似性检索特征主要有:

(1)颜色相似性。颜色是图像最重要的特征之一,现在已成为基于内容相似性检索的首选特征^[13,14]。颜色特征的提取与检索通常有两种方法:颜色直方图方法(color histogram method)和颜色对方法(color pair method)。颜色直方图又称灰度直方图,可用函数表示,对每一灰度,都对应着该灰度的图像中的像素(pixels)数,这一组数就是颜色直方图。使用该法检索时,将示例图像直方图与图像数据库图像直方图进行相似性匹配,依据一相似性公式进行计算,值越小说明相似性程度越高,越符合用户需求。该方法适用于图像的粗略查询,较细致的查询就要使用颜色对方法。颜色对方法就是借助图像子块之间颜色的邻接关系,通过颜色进行组对建模,如两幅图像具有相同的颜色对,那么这两幅图像就是相似的。

颜色相似性特征的基于内容检索的重要困难^[15]是用户的检索要求并非只有颜色条件,还伴随纹理、形状、空间等限定条件。

(2)纹理相似性。纹理就是图像局部不规则,而宏观有规律的特性。纹理特征包括粗糙性、规则性、线条相似性、凹凸性、方向性和对比度,这些特征都可作为检索项。纹理特征可使用统计方法和结构方法进行分析。

(3)形状相似性。形状特征有区域、主轴方向、矩、偏心率、圆形率、正切角等。用户可选择某一或多种形状特征,与图像进行匹配检索。基于内容的图像检索就是查询和比较查询图像和数据库图像的特征向量。

其他特征如轮廓、空间约束、领域概念等^[16],在此不赘述。值得一提的是,仅仅基于某一媒体内容特征检索往往未必能得到较好结果。因此,目前许多研究者正试图研究综合多种媒体特征合成的基于内容检索算法^[17]。

3 基于内容的视频表示和检索技术

所谓视频,又称动态图像,是一组图像按

时间的有序连续表现,视频的表示与图像序列、时间关系有关。视频的检索要比图像复杂得多,它不仅和空间,而且和时间有关。视频数据可用幕、场景(scene)、镜头(shot)、帧(frame)等描述。视频序列主要由镜头组成;镜头由一系列连续的帧组成;帧是一幅静态的图像,是组成视频的最小单位;场景含多个镜头;幕是由一系列相关的场景组成,表达一个完整的事件。

如何表示视频信息以高效检索是一个公开的难题。最流行的方法就是使用结构化建模方法^[18]表示媒体信息:视频序列首先被分割为镜头,一个镜头被用作一个基本的操作单元,表示一个原始的概念,存储在镜头级的信息包括镜头的边界和题目,伴随其他内容和上下文信息比如日期、位置、描述语等。该方法的缺点是用户难以检索视频的片断等。另一种方法是利用分层方法组织视频信息,它并非用独立的镜头表示视频信息而是把视频序列分割为一系列重叠的分层。每个分层包括题目、关键词等描述信息,一个分层可包含在其他分层中,包含许多其他的描述符。使用该方法,每个分层在逻辑上是独立的。两法的共同缺点是揭示视频信息内容的深度不够。

视频的基于内容检索的重要步骤就是视频序列的镜头分割。镜头的类型有^[19]:超长镜头(ELS)、长镜头(LS)、深焦镜头(DFS)、中长镜头(MLS)、中镜头(MS)、短镜头(CLS)、超短镜头(ECS)。一系列帧组成一个镜头,因为它们描述了相同的场景,表示了单个摄像机操作。对镜头分割的方法有两种^[20]:一是基于像颜色-密度直方图的全局表示法,如果是同一镜头,则各帧间特征直方图差异不大,如果镜头发生转换,则两镜头帧特征直方图发生较大差异,通过比较特征直方图可实现镜头分割;二是基于像密度等空间测度差异实现镜头分割,它对运动的目标和摄像机十分敏感。

镜头分割完成就确定了每个镜头的开始和结束位置以及连接镜头确定视频中每个场

景的开始和结束位置。每个镜头中的一系列帧间差异很小,可从中选择出关键帧,又称代表帧(representation frame,简称r-frame),它描述镜头的关键图像,对该镜头的存储和检索可通过对该代表帧的存储和检索实现。对代表帧的表示和检索方法与图像的基于内容表示和检索方法相同。

另外,基于摄像机和目标的运作特征也可作为检索的依据。但目前视频检索的实现在很大程度上仍依赖人工介入,自动化和半自动化程度尚不高,专家们正努力攻关,加快这一进程。

除图像、视频外,多媒体数据库还有一重要组成媒体:音频。音频媒体的语义内容也十分丰富,它的基于内容检索需借鉴语音识别的研究成果。由此可见,多媒体信息的基于内容检索研究综合借鉴多门相关学科的研究成果,只有把这些成果无缝地集成在一起,才能使其日臻成熟。以上只是基于内容的图像、视频、音频检索机制的简要阐述。

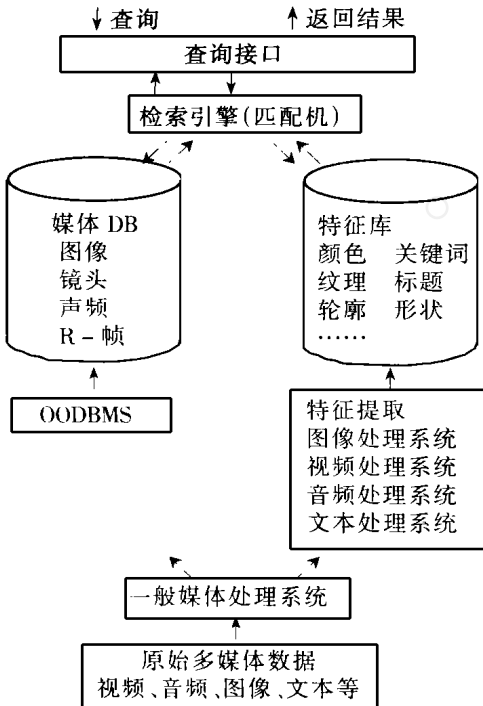
4 基于内容检索多媒体数据库的构建

现在商品化的多媒体数据中多媒体信息检索机制大多采用赋予一段视频、音频或一幅图像关键词、标题或注释等作为检索标志的方式。由于技术条件的限制,只有少数如IBM公司的关系数据库软件DB2部分地使用了基于内容检索。基于内容检索系统可作为一个检索模块嵌入到数据库系统中,实现其基于内容检索功能。

多媒体数据库系统不仅是处理音频、视频、图像,而且还有文本,是多种媒体的综合处理,它的检索机制是包括文本检索技术在内的多种检索技术的有效结合。本节在参考多种基于内容检索系统VISION^[21]、QBIC^[22]、JACOB^[23]、CHABOT^[24]等的基础上,提出了一个基于内容检索多媒体数据库的总体结构,并分别就各组成模块的功能进行阐述。

4.1 基于内容检索多媒体数据库的总体结构

基于内容检索多媒体数据库由以下 5 个子系统(或称模块)组成:一般媒体处理系统;面向对象(OO)多媒体数据库存储部分;媒体特征提取系统;检索引擎(匹配机);用户查询接口(见图 1)。



4.2 一般媒体处理系统

该子系统是整个系统的第一个处理功能模块,对输入的原始多媒体数据作预处理。包括对多媒体数据的数字化、压缩、节断化,对视频序列的镜头分割、代表帧的选取等,将预处理后的多媒体数据送到下一个子系统。

4.3 面向对象多媒体数据库存储管理系统

该子系统用于对原始的多媒体数据的存储管理,实现软件是面向对象的多媒体数据库管理系统(OOMMDBS)。目前商业化数据库管理系统多为关系型的,无法对多媒体数据进行 OO 建模,OO 方法超强的建模能力(抽象、信息隐蔽、继承等)难以发挥,因此关系型 DBMS 可舍弃不用,而采取成熟的 OO 多媒体数据模型(如 MOAP 模型、改进的 MOAP 模

型等),使用 VC++、MS-WINDOWS 等软件工具开发面向对象的存储管理器,实现对视频、音频、图像、文本等多种媒体的存储。这里存储的视频等多媒体数据是经过“一般媒体处理系统”模块处理过、视频序列已实现了镜头的分割、代表帧(R-帧)已确定,就是说数据库中存储的已是经过处理的镜头、代表帧等。

4.4 媒体特征提取系统

该子系统由图像处理系统、视频处理系统、音频处理系统、文本处理系统、领域知识专家系统等组成,其功能是自动或半自动地(用户可交互)提取媒体特征,提取用户感兴趣的、适合检索要求的特征。这些特征可以是全局性的,如整幅图像和视频镜头,也可以是针对某个目标的,如图像中的子区域、视频中的运动对象等。

图像处理系统。主要是针对静态图像,经对象识别,提取出图像的颜色、纹理、轮廓、形状、位置、空间等特征。

视频处理系统。通过镜头抽取器产生出:①代表帧;②基于运动的目标;③镜头。代表帧是一种静态图像,它可送入图像处理系统,提取颜色、轮廓等媒体特征;基于运动目标和镜头则由该系统抽取运动目标、运动摄像机视频特征。

文本处理系统是由于图像、视频等媒体中包含大量文本信息,所以要通过该系统抽取可作为检索标志的字、词、短语,以及基于上下文的检索入口等。

音频处理系统。提取音频媒体特征。

领域专家(知识)系统。针对某一具体领域的应用(如医学图像、卫星图像等),专家系统可提供领域知识辅助提取媒体特征。

所有提取的媒体特征存入媒体特征数据库,并与媒体建立索引联系,以备特征匹配使用。

4.5 检索引擎(匹配机)

该系统可把从用户查询接口获得的用户检索要求与媒体特征库里存储的媒体特征进行相似性匹配,对匹配结果进行排序,选择出

符合用户要求的媒体内容。

4.6 用户查询接口

用户查询接口的友好程度是整个系统成功与否的关键问题之一^[25]。用户接口必须设计得使用户易于选取基于内容特征,允许这些特征相互关联并与文本及参数化数据联系起来,使用户可以表述查询、一般性浏览数据库。多媒体信息查询的方式主要有两类:一是直接查询(Direct Query);二是基于示例的查询(Query By Example 简称 QBE)。所以,用户查询接口也有两种类型:字符型接口与示例型接口。

字符数值型接口主要是利用文本检索技术的成果,用字符或数值更好地体现多媒体数据库的内容以及表达用户的查询需求,字符值能否恰当地表示媒体的语义是决定性因素。该类查询接口有以下几种类型^[26]:表示类查询;关键字描述;自然语言查询。

示例查询,用户给出一个示例(可以是声音、图像等),使系统自动获取示例的特征,然后根据这些特征与多媒体数据库中的媒体特征库、媒体 DB 进行相似性匹配,获取检索结果。QBIC 系统^[27]、CHABOT 系统^[28]、JACOBS 系统^[29]等都广泛采用了 QBE 接口。有文本示例、图像示例、声音示例、视频示例、结构示例、混合示例等多种的示例接口。

用户检索接口的开发要结合使用上述两种接口类型,更好地满足用户的检索需求。

一般媒体处理系统、多媒体数据存储、媒体特征提取系统、检索引擎(匹配机)、用户查询接口等 5 大功能模块的有机结合就完成了—一个基于内容检索多媒体系统的构建。

参考文献

- 1,6 V. N. Gudivada & V. Raghavn. Modeling and retrieving images by content Information Processing & Management, Vo133 No/ 4, pp427~452, 1997
- 2,13,25,26 胡晓峰等. 多媒体系统. 北京:人民邮电

出版社,1997

- 3 李国辉等. 基于内容检索的多媒体数据查询和检索. 小型微型计算机系统,1998. 4
- 4,11,16 V. N. Gudivada & V. Raghavan. Content - Based image retrieval system, IEEE Computer, 1995, 9
- 6 J. K. Wu et al. Content - Based Retrieval for Trademark Registration. Multimedia Tools and Applications, 3, 245~267 (1996)
- 7,21 S. Guch et al. The VISION Digital VIDEO Library. Information processing & Management, vo133, No. 4, pp413~426. 1997
- 8 N. C. Rowe et al. Automatic caption localization for photographs on World Wide Web pages, Information Processing & Management, vo134. No. 1, pp95 ~ 107, 1998
- 9 Hironobu Abe et al. Content - based Management of Video in a Multimedia Authoring Environment. Multimedia Tools and Application 2, 199~214 (1996)
- 10,14,15 Philippe Aigrain, H. J. Zhang et al. Content - based Representation and Retrieval of Visual Media: A State - of - the - Art Review. Multimedia Tools & Applications 3, 179~202 (1996)
- 12 Dario Lucarella et al. A Visual Retrieval Environment for Hypermedia Information Systems. ACM Transactions on Information Systems. Vol. 14. No. 1, 1996. pp3~29
- 17 B. M. Mehtre et al. Content - based Image retrieval using a composite color - shape approach. Information Processing & Management Vol. 34. No. 1, pp109~120. 1998
- 18,19 Tat - seng Chua et al. A Video retrieval and sequencing system. ACM Transactions on Information Systems Vol. 13. No. 4. 1995. pp373~407
- 20,22,27 M. Flickner et al. Query by Image and Video Content: The QBIC system. IEEE Computer, 1995. 9. pp23~32
- 23,29 E. Ardizzone et al. Automatic Video Database Indexing and Retrieval. Multimedia Tools and Applications 4, 29~56 (1997)
- 24,28 V. E. Ogle. Chabot: Retrieval from a Relational Database of Images. IEEE Computer 1995. 9. pp40~48

蔡巍 南开大学信息资源管理系研究生。研究方向:多媒体数据库。通讯地址:天津市。邮编 300071。

刘玉照 南开大学信息资源管理系副教授,长期从事现代信息技术教学工作。通讯地址同上。

(来稿时间:1999-07-08。编发者:翟凤岐)