

赵一丹

论数字图书馆基于内容的多媒体数据 查询和检索技术

摘要 数字图书馆基于内容的检索(CBR),是指根据媒体对象的语义和上下文联系进行检索,它具有从媒体内容中提取信息检索和大型数据库(集)的快速检索等特征。基于内容的多媒体数据查询主要提供两种形式的查询说明:示例查询说明和表格说明。图1。参考文献6。

关键词 数字图书馆 基于内容的检索 基于内容的数据查询 多媒体

分类号 G250.76

ABSTRACT CBR (Content-Based Retrieval) is the retrieval according to the meanings and contexts of media objects, with the characteristics of retrieval of media contents and fast retrieval of large-scale databases. CBR provides two kinds of query statements. 1 fig. 6 refs.

KEY WORDS Digital library. Content-based retrieval. Content-based data query. Multimedia.

CLASS NUMBER G250.76

所谓数字图书馆中基于内容的检索(CBR, Content-Based Retrieval),是指根据媒体对象的语义和上下文联系进行检索,它有如下特点:

(1)从媒体内容中提取信息线索。基于内容的检索突破了传统的基于表达式检索的局限,它直接对图像、视频、音频进行分析,抽取特征。利用这些内容特征建立索引进行检索。在检索过程中,它采用相似性匹配的方法逐步求精来获得查询的结果。这一点与常规数据库检索的精确匹配方法有明显不同。

(2)大型数据库(集)的快速检索。在多媒体数据库(集)中,不仅数据量巨大,而且种类和数量繁多,因此要求CBR技术也能像常规的信息检索技术一样,快速地实现对多媒体信息的检索。

基于内容的检索要利用图像处理、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不仅仅是基于内容,而且是一门信息检索技术。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,导致新的媒体数据的表示和数据模型、有效和可靠的查询处理算法、智能查询接口、以及与领域无关的检索技术和系统结构的产生。CBR与模式识别、图像理解、计算机视觉等学科的重要区别是,CBR是一种信息检索技术,要能够从

大型分布数据库(集)中以用户可以接受的响应时间查询到要求的信息,它不需要去理解或识别媒体中的目标,关注的是基于内容快速发现信息。

1 基于内容查询和检索的系统结构

1.1 系统构成

作为一个信息检索的核心技术, CBR应接入或嵌入到其他多媒体系统中去,如超媒体(浏览器)系统、多媒体信息系统、关系数据库系统等,提供基于多媒体数据内容的信息查询和检索。因此,把CBR设计为多媒体数据库(信息)进行存取的检索引擎结构。CBR系统一般由两个子系统构成:数据库生成子系统和数据库查询子系统。每个子系统由相应的功能模块和部件组成,如图1所示。

1.2 基于内容的查询和检索过程

基于内容的查询和检索是一个逐步求精的过程,其检索存在一个特征调整、重新匹配的循环过程。其过程如下:

(1)初始查询说明。用户查找一个对象时,最初可以用QBE或查询语言来形成一个查询。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的特征按照一定的匹配算法进行相似匹配。

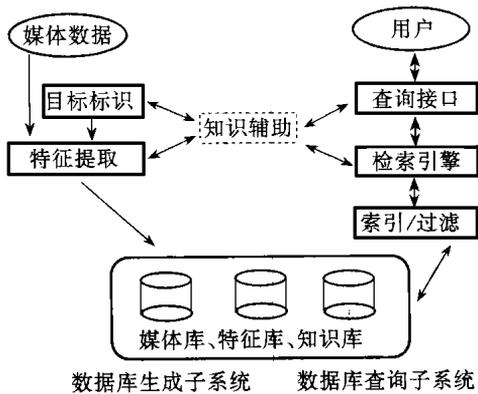


图 1 CBR 的系统结构

(3) 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。

(4) 特征调整。对系统返回一组满足初始特征的查询结果,用户可以通过遍历(浏览)来挑选到满意的结果,或者从候选结果中选择一个示例,经特征调整,然后形成一个新的查询。

(5) 如此逐步缩小查询范围,直到用户满意为止。

2 用户接口

2.1 查询和检索分类

基于内容的检索根据媒体对象的不同,可分为基于图像、视频、音频、文本、图形等媒体内容的检索。

按媒体的时空关系分类,可分为按时序关系和按空间关系的内容检索。按时序关系分,有基于音频旋律、视频镜头操作、运动目标的检索等;按空间关系分,有基于图形中的图元、图像中的目标之间的空间关系的检索等。

按照提取的特征的性质不同,可分为:基于媒体物理(原始)特征检索(如颜色、纹理特征等)和基于媒体逻辑特征(人工或半自动提取的特征,目标识别特征等)的检索。

2.2 查询说明的形式

无论是用哪种形式来形成一个查询,都应以最方便的图形查询接口方式向用户提供,即在图形环境下,提供数据库查询的一致视图。CBR 主要提供以下两种形式的查询说明。

(1) 示例查询说明(QBE)。用户通过一个例子来表达查询的要求。这个出示的例子可以是用户选择(系统提供的示例模板)或绘制的,也可以是通过浏览,选择库中的某个媒体作为例子。颜色、纹理、形

状、草图均可以作为示例进行查询,系统从用户的示例中实时提取特征用于检索。

(2) 表格说明。用户往往难以描述一种较为复杂的查询要求,这种情况下,向用户提供一致的表格形式来形成文本、音频的查询、运动查询,以及利用可视 SQL 语言进行查询。表格中的每个描述域包括项名,以及可能的描述值。CBR 将利用用户的主客观属性,以及自动预提取的特征进行检索。

3 快速存取方法

当数据库中包含成千上万的图像、视频、文本等媒体数据时,为了有效地查询,合适的存取访问结构和检索方法就变得非常重要。传统的索引方法已不再适用,因此需要研究新的适合于多媒体内容特征匹配的快速访问结构。对于大容量媒体库的快速搜索,可以用以下两种技术。

3.1 索引

索引是对特征库的快速访问。对于数据库中每个数据项(特征项),索引项包含关键属性值,以及可以直接访问数据项的指针。索引构成树结构。一个索引树既可以自底向上通过抽象来构造,也可以自顶向下通过分类来构造。

多媒体数据的索引树,是利用一种抽象数据类型,它可以是特征(特征矢量)、多维矩阵(多维特征、图像数据、图像序列等)或指向数据结构的指针。在索引树的不同级别上,允许所用到的关键属性可以不同。例如,在索引树的顶层,关键属性可以是面部轮廓特征,在下一级可以是头发特征。另外,分组的准则和函数不仅是基本的逻辑表示式,例如, 01 ai 02 ,而且主要是特征矢量的相似性度量,或用集束(cluster)的方法来归类相似的视频镜头,还可以利用特征的空间、时间约束来分类等。

从宏观上看,索引可以分级来加快数据访问。索引级最高层是最概括的总目,下级是逐步缩小范围的具体索引项。

索引可以分类,以向用户提供不同的检索方法。例如,在一个视频数据库中,可以提供 3 种类型的索引:目录索引(标题、类别等);结构索引(镜头、场景等);内容索引(场景中的角色、运动目标等)。

3.2 过滤

用快速计算的过滤器扫描数据库中的所有特征数据,只有通过过滤器的项才完整地计算相似度,这样可以加快检索过程。每种查询类型,都有相应的过

滤波器。例如,对于颜色查询,首先用非常快的三维颜色空间算法对1万个元素的数据库进行一次过滤,产生1000个初步的匹配集合,然后这些已过滤的元素顺序进入较慢的256维颜色直方图(矩阵矢量相乘)的计算和匹配操作,最后获得最佳的20幅匹配图像显示给用户。

4 视频检索和浏览

4.1 视频数据的结构化特征

现有的多媒体信息系统,包括能够存储多媒体数据的关系数据库系统和VOD系统,对视频媒体仅仅局限于存储和关键字的检索。这样的系统所管理的数据资源是非结构化的,即没有对视频资源做任何分析,因此不能索引,不能进行基于结构特征的存取。要对视频进行真正意义上的查询和检索,就要抽取并描述视频的结构化特征。

任何视频都是由一个个镜头衔接起来的,因此镜头是视频检索的基本单元。对视频中的镜头分割是视频分析中最基本的内容。视频分割即将视频数据分割为一个个镜头的过程,其核心处理是识别镜头的切换。

镜头切换主要有突变和渐变两种:突变是指一个镜头与另一个镜头之间没有过渡,由一个镜头的瞬间直接转换到另一个镜头;渐变是指一个镜头到另一镜头的渐渐过渡过程,没有明显的镜头跳跃。渐变包括淡入和淡出、渐隐渐现、划入划出等。

视频分割成镜头后,要从每个镜头中抽取代表帧,简称R帧,又称关键帧。代表帧是用于描述一个镜头的关键图像,它反映镜头的主要内容。对代表帧的特征提取和一般静态图像的特征提取是一样的,包括颜色、纹理和形状等特征。

代表帧除具有静态特征外,还具有动态特征。视频的运动特征主要有:

(1) 摄像机操作。如摇镜头(paning)、推拉(zooming)、跟踪(tracking),以及镜头的其他操作(仰视拍摄等)。

(2) 目标运动。目标运动可以用运动方向和运动幅度来描述。事实上,许多目标的运动也与摄像机操作有关,通过对视频的研究发现,当目标运动时,在视频上表现为背景在迅速变化,运动目标实际上相对镜头没有太大的运动,而是摄像机在操作,这样就可以利用摄像机操作的特征进行检索。

4.2 视频检索和浏览

一旦建立了视频内容,就可以在这些内容上进行

基于内容的视频检索和浏览。查询过程仍然可以迭代,以系统可以接受的反馈重新形成搜索。

4.2.1 基于关键帧的检索

一旦视频被抽象为关键帧,搜索就变成按相似度来检索那些在数据库中与查询描述相似的关键帧。提供的常用查询方法是通过目标特征说明的(直接)查询和通过可视实例的(示例)查询。检索时,用户也可以指定使用特定的特征集。

一旦检索到关键帧,用户就可以利用播放来观看它所代表的视频片断。浏览可以跟随检索,检验检索到的关键帧的上下文边界联系。另一方面,浏览也可以初始化查询,即当浏览时,用户可以选择一个图像作为查询,找到所有与该图像相似的关键帧。

4.2.2 基于运动的检索

基于镜头和目标的时间特征来检索镜头是视频查询的进一步要求。如查询“找到摄像机平扫10度的所有镜头”,就可以利用摄像机操作的表示来查询。另外,用运动方向和运动幅度特征可以检索到运动的主体目标。在一个查询中可以结合时间和关键帧特征,这样可以检索出虽然具有相似的运动特征但静态的颜色特征不同的镜头。

4.2.3 浏览

除了查询和检索,对于视频来说,浏览也同样重要。通过浏览,我们可以对任何不了解的内容细细看一看。视频浏览一般采用分层结构和集束分类技术。

分层的浏览器提供对视频任何点的随机存取。视频序列在空间上散布并用代表帧图标表示,就像幻灯片。换句话说,显示空间以镜头的代表帧表示,从而提供长视频内容的快速总览和存取。

为了支持基于分类的浏览,要使用集束算法。可以用分层的集束算法。即用关键帧和镜头特征对镜头进行集束分类。每个类别由相似内容的一组镜头组成。集束分类后,每类镜头用一个图标表示,然后显示在分层浏览器的高层上。结果,用户就可以粗略知道每个镜头的内容,而不需要进入下个层次。

5 需要继续探讨的问题

基于内容的查询和检索研究已取得许多成果,推出了一些原型系统,某些技术已试用于国内外一些数字图书馆。然而还有许多问题需要探索和研究。

(1) 数据模型。数据模型是CBR系统的核心。模型决定CBR支持的查询类型和检索能力。实现的

(下转第76页)

于过渡和转换阶段。新的高校图书馆运行机制模式应当怎样?从美国高校图书馆运行机制的研究中,我们可以得到一些启示。

打破图书馆管理上的条块分割。我国高校图书馆的管理体制,基本上是封闭的、平衡的结构。在政府—学校—财务部门—政府这一链条中间缺少市场要素和参与,缺乏可调节性与灵活性,难以适应外界尤其是知识经济改革浪潮与网络化时代教育的需要。

引入市场竞争机制的同时,强调政府的立法职能与宏观调控。政府除宏观调控外,将其余各项职能转向社会,这样,高校图书馆的经费问题在很大程度上可以因为引入了社会资源,与社会的合作而得到缓解。

突破只服务于本校师生的狭窄经营理念。在网络化浪潮遍及全球的今天,实现教育知识与资源的共

享已是高校图书馆的一个基本方向。只有实现馆际间的资源共享,才有可能为用户提供最好的服务,才能赢得市场,更好地发展。

参考文献

- 1 A Guide To Cornell University Library 1999~2000
- 2 易丹柯.论高校图书馆信息业的发展.中国图书馆学报,1999(4)
- 3 赵达雄.中外图书馆开展信息服务之比较研究.图书馆理论与实践,1998(1)
- 4 孙媛.提高综合效益实现管理创新——新建浙江大学图书馆的探索与思考.大学图书馆学报,1999(5)

薛小棉 浙江大学教育系资料室馆员。通讯地址:杭州市。邮编 310028。

(来稿时间:2000-09-18)

(上接第 59 页)

数据模型,需要具有领域无关的和有效存储的物理特征和逻辑特征。

(2) 查询说明模式。对于特定的查询,领域专家可能会需要一种功能丰富的图形接口,以支持多种形式的查询类型。而数字图书馆的普通用户可能更愿意用自然语言来说明查询。映射自然语言查询到通用查询类型仍然需要进一步研究。

(3) 分布。随着信息高速公路的出现,Web 网上存在图像、视频等媒体仓库。必须研究能够提供快速、透明地访问这种分布多媒体数据集的方法。

(4) 特征提取和语义获取。需要容易使用的抽取原始特征的自动工具和获取逻辑特征的半自动工具,用于开发大型、实用的多媒体检索应用。另外还需要新的如基于分形或小波的特征数学表示。

(5) 性能。研究适合与大型媒体库检索的索引结构和过滤器,缩短系统响应时间,提高效率因子。

(6) 集成查询。系统需要能够把日期、价格等客观属性与内容特征如颜色、纹理和形状等结合在一起的集成查询。

(7) 扩展性和灵活性。系统体系结构应该支持更新和扩充新特征和新匹配—相似性度量算法,以使系

统查询和检索的性能和功能不断增强。现在我们使用的媒体格式和编码没有考虑到它包含的内容,只是针对像素和样值来编码,因此从这些数据中抽取内容特征非常困难。如果我们在对媒体数据编码表示时就考虑到媒体的内容,那么对这些数据的内容检索就会更有效和准确。

参考文献

- 1 张丽虹.数字图书馆及其相关问题和技术研究.图书馆杂志,1997(1):32~34
- 2 汪冰.数字图书馆:定义、影响和相关问题.中国图书馆学报,1998(6):9~17
- 3 胡晓峰,李国辉.多媒体系统.北京:人民邮电出版社,1997
- 4 K. Klemperer, et al. Digital libraries: a selected resources guide. Information Technology & use 1997, 16(3):126~131
- 5 Andrew Wang. Library information service in the 21st century. In: ISAL '96, Shanghai, Sept:1~4
- 6 Dr. Frank Tung. "IBM Digital Library" Lecture, 1997

赵一丹 女 武汉大学信息学院传播学硕士生。通讯地址:武汉市。邮编 430072。

(来稿时间:2000-10-23)