基于 MPEG-7 的多媒体搜索引擎构建*

范新华, 陈宏兵, 许满武
(南京大学 软件新技术国家重点实验室, 江苏 南京 210093)

摘　要: 介绍了多媒体搜索引擎的工作原理, 叙述了搜索引擎的系统结构以及各模块的具体实现, 最后讨论了 MPEG-7 搜索引擎的特点和潜在应用价值。

关键词: MPEG-7; 多媒体搜索引擎; 特征描述子
中图法分类号: TP37　　**文献标识码**: A　　**文章编号**: 1001-3695(2004)11-0187-04

Construct a Multimedia Search Engine Based on MPEG-7

FAN Xin-hua, CHEN Hong-bing, XU Man-wu
(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing Jiangsu 210093, China)

Abstract: Introduces the working principle of multimedia search engine, and brings forward its architecture and concrete realization of each part's functions. Finally discusses the character and potential application value of the search engine based on MPEG-7.

Key words: MPEG-7; Multimedia Search Engine; Feature Descriptor

在现代信息社会中, 上网查找所需的资料已成为人们获取信息的重要手段之一。随着多媒体技术和 Internet 技术的飞速发展, 网络上可用的多媒体信息急剧增加, 这些信息以多种格式存在, 主要有图像、音/视频等。随之人们对使用搜索引擎工具检索多媒体信息提出了越来越高的要求。目前 Internet 上提供的多媒体搜索引擎大多采用文本检索方式, 如 Lycos, HotBot, Google, Yahoo, AltaVista 等。但是, 图像、视频包含的信息量丰富, 文本描述有时不能恰如其分地反映其中的内容, 用户也不能对查询要求用文字合理描述, 导致检索出的结果往往不能很好地符合用户的需求。鉴于此, 基于内容的检索技术已成为人们研究的热点之一, 其主要方法是依据多媒体内容的底层特征作为多媒体信息的索引, 计算查询示例和目标媒体对象的特征相似距离, 按相似度匹配进行检索。由于基于内容的检索有着广泛的需要, 并有着较好的市场前景, 因而也引起了国际标准化组织的关注。继 MPEG-1, MPEG-2 和 MPEG-4 之后 MPEG(Moving Picture Experts Group, 移动图像专家组) 又制定了一个国际标准 MPEG-7。MPEG-7 对各种不同类型的多媒体信息进行了标准化描述, 并将该描述与所描述的内容相联系, 以实现快速有效的搜索。MPEG-7 只标准化了多媒体信息的内容描述, 而描述的生成(包括特征抽取、索引建立) 与描述的使用(包括搜索引擎、检索处理) 都不在 MPEG-7 标准之内, 目的是为了 MPEG-7 能够更广泛地被应用; 而且引入了竞争机制, 能够针对不同的应用领域产生更好的特征抽取算法和搜索引擎。MPEG-7 的出现, 将会大大地促进多媒体信息的分析、存储、检索、过滤等领域的研究, 同时也使得人们可以像目前在互联网上检索文本信息一样检索多媒体信息。

收稿日期: 2003-11-10; 修返日期: 2004-03-02
基金项目: 国家自然科学基金资助项目(60273035); 国家"863"计划资助项目(2001AA113161); 江苏省自然科学基金资助项目(BK2002080)

1　MPEG-7 标准概述

MPEG-7 是由 MPEG 制定的一个 ISO/IEC 标准。MPEG-7 正式名称是多媒体内容描述接口, 目的在于为描述多媒体内容提供一个统一标准。MPEG-7 描述的音/视频数据包括静态图片、图形、3D 模型、语音、音频、视频以及关于这些元素如何组合在多媒体表示中的合成信息。MPEG-7 允许不同的描述粒度, 可以提供不同级别的辨别能力。根据应用领域的不同, 相同的内容可以使用不同类型的特征描述, 如对于视频信息来说, 较低的抽象级别是对形状、纹理、颜色、移动和位置等方面的描述, 而高层次的描述则给出了语义方面的信息。

MPEG-7 标准化的主要元素有描述子(Descriptors, Ds)、描述模式(Description Schemes, DSs)、描述定义语言(Description Definition Language, DDL) 和一系列的系统工具。描述子定义了每个特征表述的语法和语义, 主要用于底层的视听特征, 如颜色、纹理、运动和声强。描述模式可以包含描述子或者其他描述模式, 详细说明其组件之间的结构和语义; 描述模式着重在某些重要的媒体信息概念上, 如人物、视频片断、层次概述等。描述定义语言用于创建新的描述子和描述模式, 以及允许对存在的描述模式进行扩展和修改。MPEG-7 的系统工具用于支持描述的复合、内容同步和传输机制等。

本文构建的搜索引擎采用的是 MPEG-7 标准中的视觉描述工具。在 MPEG-7 标准中, 视觉描述部分定义了视觉数据信息的描述子和描述模式, 包括颜色、纹理、形状、运动、位置等共六类基本视觉特征描述, 每个方面的特征又包括多个不同的描述子, 描述媒体某一具体的方面。其中, 颜色的主要描述子有颜色空间、主颜色、颜色直方图、颜色量化、GoF/GoP 颜色、颜色结构、颜色分布及可扩展颜色等; 纹理的主要描述子有边缘

直方图、相似纹理及纹理浏览; 形状的主要描述子有基于区域的形状、基于轮廓的形状及 3D 形状; 运动的主要描述子有摄像机运动、目标运动轨迹及参数目标运动和运动活度; 位置的主要描述子有区域定位器和时空定位器; 还有其他用于人脸识别的描述子。

2 搜索引擎的设计

2.1 工作原理

对于多媒体信息的检索, 互联网上现有的搜索引擎大多都是基于文本描述的, 即对多媒体信息添加文本说明。这种方法主要是对含有多媒体信息的网站和网页进行分析, 对多媒体信息的物理特征和内容特征进行著录和标引, 把它们转换成文本信息或者添加文本说明, 建立标注数据库, 检索时主要在此数据库中进行布尔匹配。一般说来, 可以用于检索的文本信息有: ①文件扩展名和超文本标志; ②标题和文字描述; ③人工标引的信息。

本文中构建的多媒体搜索引擎主要是针对图像/视频媒体的检索和获取。对于图像, 直接提取图像的视觉特征来建立图像的索引; 对于视频, 首先从整段视频中切分出镜头后, 再从镜头中抽取关键帧, 并通过提取这些帧的底层视觉特征来建立视频的索引。MPEG-7 搜索引擎主要采用的是示例图像的检索方式, 即用户给出一幅类似的图像、手绘图像的轮廓、大致的对象形状等, 根据用户给出的示例, 系统按要求得到相应的特征, 然后对数据库进行匹配查询, 最后将得到的结果按相似的程度返回给用户。对于一幅图像而言, 它主要具有颜色、纹理和形状这三个方面的视觉特征。MPEG-7 标准视觉部分中为每个方面的特征规定了多种特征描述子, 每种特征描述子都可以作为搜索引擎的一种检索方法。我们在这三个方面各选取一个特征描述子分别进行特征抽取和匹配, 并将三个特征描述子的相似性匹配度按照一定的权重来形成总的相似性匹配度, 这样可提高搜索的准确性; 同时用户也可以根据自己的需要选择某一特征进行检索。本系统中选择的三个描述子分别是: 可扩展的颜色、相似性纹理和基于轮廓的形状。另外系统也提供了基于关键字的检索方式, 用户可以首先输入关键字进行检索, 然后在返回的结果中选择图像或关键帧, 根据相关内容特征进行进一步的检索。

2.2 引擎客户端软件的功能需求

客户端软件是用户访问搜索引擎的接口, 表现出搜索引擎实现的各种功能。通过客户端软件, 用户能够:

(1) 选择检索方式。如果选择采用关键字进行检索, 用户输入关键字; 如果选择示例检索, 用户输入检索的源图像或者选择浏览到的图片。

(2) 选择检索方法。列出搜索引擎支持的所有的检索方法, 用户可选择某一种方法进行检索。

(3) 控制检索结果数量。用户需输入最多返回多少个检索结果, 引擎将丢弃超过的检索到的媒体信息。

(4) 显示检索结果。搜索引擎按照匹配度对检索结果进行排序, 客户端显示排序后的检索结果。

(5) 访问检索到的媒体。如果是图像, 可直接显示给用户; 如果是视频媒体, 可下载或者用缺省的播放器进行播放。

2.3 搜索引擎的系统结构

整个系统结构包括搜索引擎服务器、多媒体著录系统、多媒体数据库服务器、Web 服务器以及检索客户端等几个部分组成。检索客户端又分成本地客户和通过 Web 访问的远程客户。各模块的功能如下:

(1) 搜索引擎服务器。该服务器负责接收来自本地客户机或者 Web 服务器的检索请求, 读取数据库中的特征数据进行源特征和目标特征的匹配, 按照匹配度返回给客户端检索结果。

(2) 多媒体著录系统。对各种需要著录的媒体按照选定的特征抽取方法, 生成相应的特征数据, 并将特征数据按照抽取的方法分类存放到数据库中, 同时将特征数据和媒体的位置关联起来。

(3) 数据库服务器。保存媒体抽取生成的特征数据, 为搜索引擎提供高速的特征读取服务。

(4) Web 服务器。以 Web 服务的方式代理远程客户的检索服务。

(5) 本地客户。供本地用户直接访问搜索引擎进行检索, 安装有本地客户端软件, 用户通过该软件访问搜索引擎。

(6) 远程客户。Web 用户可通过浏览器下载检索客户端软件, 通过该软件来进行搜索服务。

2.4 多媒体数据库的设计

MPEG-7 标准为描述子特征值的存放给定了两种方式: ①XML 格式存放的文本形式, 便于对特征进行编辑、查找和过滤; ②一个二进制方式存放的 BiM(Binary of MPEG-7) 形式, 便于特征的存储、传输和分发。在我们的系统中, 抽取出来的特征值统一存放在数据库的表格中, 便于进行搜索引擎查询。与其他 MPEG-7 系统进行数据交换时, 则用相应的转换程序, 将以数据库表格形式存放的特征值转换为 XML 文本方式或者 BiM 二进制格式。数据库的设计是按照一个搜索方法一个表格的方式来组织的, 每个搜索方法对数据库的访问都只涉及到一张表, 这样可以获得较好的系统性能。目前我们选择的搜索方法表定义如下:

```
SC_HT_CS( nterId int not null PRIMARY KEY,  //序号
          scvalue BLOB,  //可扩展颜色描述子的特征值
          htvalue BLOB,  //相似纹理描述子的特征值
          csvalue BLOB,  //基于轮廓的形状描述子的特征值
          url varchar(255),  //多媒体的 URL
          description varchar(255))  //简单描述
```

3 搜索引擎系统的实现

我们的目标首先是实现一个最小功能的、可使用的搜索引擎系统, 包括 Web 客户端、Web 服务器、搜索引擎服务器、多媒体著录系统和数据库系统等部分。在搜索方法上也是首先采用一种综合可扩展颜色、相似纹理、基于轮廓的形状三种特征描述子的检索方法, 在此基础上可再添加另外的检索方法, 提供更好的检索策略。在实现时充分考虑了系统的可扩展性。

3.1 特征描述子的实现

采用面向对象的技术来实现各个特征描述子。将特征描述子实现为 C#的一个类, 封装该特征描述子的各种方法和属性值, 主要的方法有特征值的抽取、特征值的匹配等。我们从 MPEG-7 标准视觉部分的颜色、纹理和形状特征中分别选取了

可扩展颜色描述子、相似纹理描述子和基于轮廓的形状描述子, 进行视频关键帧和图像特征的抽取与匹配。下面介绍各特征描述子的抽取和匹配算法。

(1) 可扩展颜色描述子。该描述子是 HSV 颜色空间的一个颜色柱状体, 使用 Haar 变换编码。它的二进制表示在 Bin 的数量上是可以扩展的(从 16 位到 256 位), 位表示的精确度根据数据速率的不同也是可以变化的。

抽取方法　特征的抽取包括 HSV 颜色空间的一个柱状体的抽取(根据标准统一量化成 256 个 Bin) 和柱状体值的抽取, 被非线性地量化。

匹配方法　从 Haar 系数来重新构建颜色柱状体使得匹配有着最高的检索效率。通过 Haar 系数域的匹配来计算 Hamming 距离, 它的复杂度非常低。Hamming 距离通过比较两个 63 位的描述子和通过找到不同二进制位位置的数量来计算。实现此方法的一种方式是首先计算两个描述子的异或值, 再计算这些异或值的位为 1 的数量。

(2) 相似纹理描述子。纹理表示图像的分布和序列性, 如方向性、粗糙、式样的匀称等。相似纹理描述子为纹理图像数据库提供基于 Image- to- Image 的相似性。为了描述图像的纹理, 图像的能量和能量偏差被从频率分布中抽取出来, 用于组成基于相似性检索的相似纹理向量。

抽取方法　该描述子的第一和第二部分通过计算图像像素强度的平均值和标准偏差来抽取, 其余的能量和能量偏差通过使用频率域内 30 个 Gabor 滤波器来计算(六个方向和五个尺度), 如图 1 所示。具体算法如下:

①图像纹理的 Radon 变换(空间域分布的统计), 再进行 Fourier 变换(空间域分布转换为频率域分布), 变换后的结果为 $P(\omega, \theta)$。

②用 Gabor 滤波器滤波 $P(\omega, \theta)$。

$$H_i(\omega, \theta) = G_{P_{s,r}}(\omega, \theta) \cdot |\omega| \cdot F(\omega, \theta)$$

其中, $i(=6 \times s + r + 1)$ 为特征通道索引, s 和 r 分别表示 s^{th} 尺度和 r^{th} 方向。

③通过以下公式计算能量 e_i 和能量偏差 d_i

$$p_i = \sum_{\omega} \sum_{\theta} H_i^2(\omega, \theta) \text{ 和 } e(i) = \log[1 + p(i)]$$

$$q_i = \sqrt{\sum_{\omega} \sum_{\theta} [H_i^2(\omega, \theta) - p_i]^2} \text{ 和 } d_i = \log[1 + q_i]$$

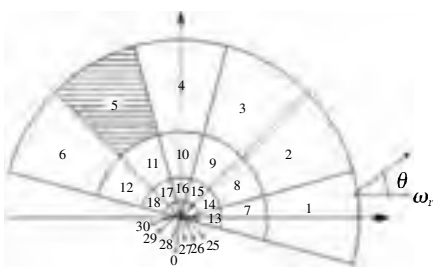

图 1　纹理抽取中的频域分布图

匹配方法　两个图像纹理的相似性距离通过计算两组特征向量的带权值的绝对差总和来度量, 一个特征向量是用于表示查询图像的纹理特征(TD_{query}), 另外一个是数据库中的图像的纹理特征($TD_{Database}$)。相似性是通过计算两个特征向量的距离来度量的, 由下式给出:

$$d(TD_{query}, TD_{Database}) = distance(TD_{query}, TD_{Database}) = \sum_k \left| \frac{TD_{query}(k) - TD_{Database}(k)}{\alpha(k)} \right|$$

其中, $\alpha(k)$ 是用于给定数据库的 $TD_{Database}(k)$ 的标准偏差系数, 但是用户可以指定他自己选择的 $\alpha(k)$ 值, 如 $\alpha(k) = 1$。

(3) 基于轮廓的形状描述子。该描述子用于描述图像或视频系列的一个 2D 对象或区域的封闭轮廓。轮廓形状描述子支持轮廓形状相似性匹配和基于轮廓形状的滤波, 这种轮廓形状的表示非常简洁, 大小在 14 字节以下。

抽取方法　本描述子从一个 2D 区域或对象的轮廓点(边界点) 列表中抽取出来, 将轮廓点的坐标转换为曲线的曲率和半径, 形成所谓的顶点图像(CSS Image)。

匹配方法　匹配通过计算查询形状描述子和数据库中的模型形状描述子的相似性度量来完成。首先比较查询和模型形状的全局参数, 如果它们存在重大的不同就没有必要进行下面的比较了。如需进一步比较, 对于剩余的形状, 相似性度量计算为全局参数之间和查询与模型形状相关的 CSS 顶点之间的相似性度量的带权重的总和。

3.2　系统各功能模块的实现

由于用于查询的示例图像文件一般都比较大, 如果客户端将待匹配的示例图像传到 Web 服务器端的中间层进行特征抽取, 会造成网络的拥塞和检索响应的延迟; 同时, 对图像进行特征抽取的计算量也比较可观, 在中间层进行抽取必将严重消耗 Web 服务器的资源, 限制了可同时提供服务的用户数。考虑到客户端和服务器端的负载平衡, 我们把示例图像的特征抽取安置在客户端进行。整个检索系统采用. NET 的框架来实现, 系统结构和基本流程如图 2 所示。

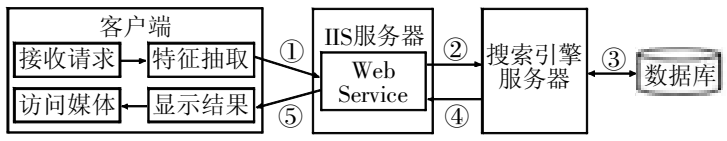

图 2　搜索模型流程图

(1) Web 客户端软件的实现。Web 客户端软件是用户使用搜索引擎的接口, 其工作流程如下: 接收用户的搜索请求; 按要求进行特征的抽取, 将请求发送到搜索服务器; 将服务器发回的结果显示给用户; 用户访问搜索到的媒体。

(2) 搜索引擎服务器的实现。搜索引擎服务器接收客户端传送来的查询请求, 根据请求中指定的搜索方法进行搜索, 并将搜索的结果回送到客户端。其工作流程如图 3 所示。

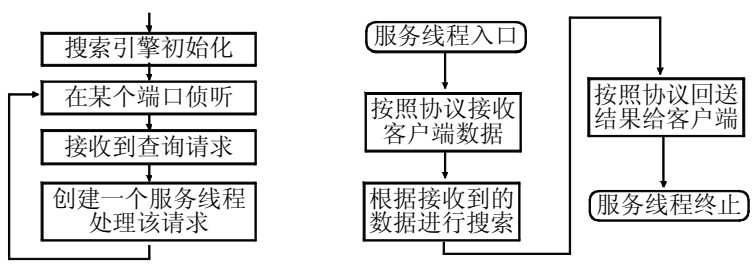

图 3　搜索引擎内部工作流程

(3) 客户端和服务器端通信协议。客户端按照如下格式向服务器发送请求: ①发送请求起始标志; ②发送选取的搜索方法(发送搜索方法的内部编号); ③发送应返回的最多结果数量; ④发送客户端抽取的特征值; ⑤发送请求结束标志。服务器端按照如下格式回送给客户端搜索结果: ①发送响应起始标志; ②发送搜索结果集合数量; ③发送每个符合要求的检索结果(包括匹配度、描述信息和 URL); ④发送响应结束标志。

(4) 其他部分的实现。多媒体著录系统提供图形界面、输入媒体的 URL 和相关描述信息, 启动特征抽取操作, 并将特征

值保存到多媒体数据库中。若是图像,多媒体著录系统可直接进行特征抽取操作;若是视频,首先要通过有关视频分析工具进行镜头分割,然后抽取有效代表镜头内容的关键帧,著录系统再对抽取出的关键帧进行特征抽取操作。

3.3 实验结果

图 4 给出了示例检索的界面。从图 4 中可以看出,综合三种特征的搜索方法取得了较好的检索结果。系统可以根据用户的需求,选择单个特征或多个特征组合,并设置各特征的权重,从而采取不同的检索策略,得到用户满意的检索结果。

图 4 示例检索界面

4 多媒体搜索引擎的特点及其潜在的应用价值

4.1 基于 MPEG-7 的搜索引擎的特点

基于 MPEG-7 的搜索引擎是按照多媒体信息的内容特征进行搜索的,用户只用给出类似的示例图像或大体的轮廓就可以查找到所需的信息。与基于文本描述的搜索引擎相比,它的主要优点有:可以自动地从多媒体信息中提取特征,避免了费时费力的人工标注;采用了 MPEG-7 标准的描述子进行特征的比较,这样使得底层的设计有着统一的基础,易于引擎之间的信息共享;搜索的结果按照特征的相似匹配度排序,用户更能获得理想的检索结果。不足之处主要有:与基于关键字的检索相比,特征值之间的匹配需要一定的时间,使得返回结果的速度比较慢。

4.2 潜在的应用价值

基于 MPEG-7 的检索方面的应用是 MPEG-7 的一种推(Pull)应用,它提供给人们一种全新的信息组织和获取的方式,具有很好的应用前景和价值。

(1) 促进数字图书馆的建设。在 20 世纪末全球信息高速公路建设热潮中,数字图书馆成为许多发达国家关注的焦点之一,这是国家信息基础设施建设不可缺少的重要内容,也是新形势下国家间资源争夺的一个新领域。数字图书馆所面临的技术问题主要有信息资源建设、存储与压缩、分类、索引和检索、安全性等,这都是亟需解决的,但其中最关键的还是统一技术标准。MPEG-7 标准完全满足数字图书馆建设的基础技术要求,MPEG-7 能够为数字图书馆保存的各种格式的信息定义标准化的描述,便于多媒体信息的检索和信息交流,同时 MPEG-7 还提供了对多媒体信息的知识产权、存取权限、数据安全管理的支持。

(2) 视频点播系统(VOD)。VOD 是能在用户需要时随时提供交互式视/音频服务的系统,在网络上具有提供给各个用户对大范围的音频节目、视频节目、多媒体信息进行同时访问的能力。然而随着音/视频节目的大量增加,用户和节目之间的互动查询变得越来越困难,我们可以应用 MPEG-7 技术来组织和管理这些节目。在视频归档时,可以自动或半自动地提取视频的特征,按照 MPEG-7 的格式描述,这样可以更方便地编辑视频节目。而对于用户来说,则可以更直观地浏览、检索视频节目。

5 结束语

本文构建了一个基于 MPEG-7 的多媒体搜索引擎,对将 MPEG-7 标准应用到基于内容的多媒体检索进行了初步的尝试。从目前的试验情况看来,视频、图像的检索取得了较好的结果。但是 MPEG-7 的特征抽取和匹配在速度上还无法和文本搜索引擎相比,因此基于 MPEG-7 的搜索引擎要在互联网上广泛地应用还需要进一步的研究。我们下一步的主要工作有:①加入更多的特征描述子,作为搜索引擎的检索方法,以此获得更好的检索结果。②在特征描述子的抽取、匹配算法中作进一步的改进,从而获得较快的检索速度。

MPEG-7 从制定以来已有一段时间了,国际、国内关于 MPEG-7 的项目逐渐开展起来,但大多尚处于研究和试验阶段,真正的基于 MPEG-7 的成熟应用还没有出现。同时,MPEG-7 是一个不断发展的标准,随着 MPEG-7 应用实践的不断深入,MPEG-7 标准将会不断地完善。对于多媒体信息的检索,基于内容的检索技术将会逐步取代当前的文本检索方式,成为人们获取多媒体信息的主要方法。

参考文献:

[1] - F Chang, T Sikora, A Puri. Overview of MPEG-7 Standard [J]. IEEE Trans. Circuits Syst. Video Technol., 2001, 11(6): 688-695.

[2] S Manjunath, et al. MPEG-7 Color and Texture Descriptors[J]. IEEE Trans. Circuits Syst. Video Technol., 2001, 11(6): 703-715.

[3] Yamada, et al. ISO/IEC JTC1/SC29/WG11/N3673, MPEG-7 Visual Part of Experimentation Model 8.0[S].

[4] homas Sikora. The MPEG-7 Visual Standard for Content Description: An Overview[J]. IEEE Trans. Circuits Syst. Video Technol., 2001, 11(6): 696-702.

[5] Brin, L Page. The Anatomy of a Large-scale Hypertextual Web Search Engine[C]. Proceedings of the 7th World Wide Web Conference, 1998. 107-117.

[6] ttp://www.medialab.sonera.fi/workspace/MultimediaSearchEngines.pdf, 2002-01[EB/OL].

[7] 单锦来, 许满武, 等. MPEG-7 和 MPEG-7 实验模型参考软件[J]. 计算机科学, 2003, (6).

作者简介:

范新华(1979-),男,硕士研究生,主要研究方向为多媒体信息检索、新型程序设计;陈宏兵(1971-),男,博士研究生,主要研究方向为多媒体信息检索;许满武(1944-),男,教授,博士生导师,主要研究方向为新型程序设计。