

基于内容检索的视频数据库系统

王志强 蔡平 (深圳大学信息工程学院 518060)

摘要: 本文首先介绍了视频数据的建模和系统设计, 讨论了视频分割、代表帧及其特征提取, 并提出了一种基于内容的视频检索和浏览方法。

关键词: 视频检索 视频分割 视频数据库

1 引言

随着多媒体和网络技术的迅速发展, 多媒体信息越来越多地出现在 Web 站点上。这就使得人们有机会接触到大量的视频数据, 因此如何有效地组织、管理和检索大规模的视频数据库将成为未来的一个研究热点。由于视频数据是一种非格式化、时空并存、数据量大的信息, 传统的基于文本的信息查询技术已不能满足人们的要求, 而基于内容的检索技术便是解决这一问题的关键技术之一。基于内容的检索技术是指基于信息的内容或特征对多媒体信息进行有效的检索。所谓信息的内容或特征主要是指视频数据中的帧、镜头、运动对象和场景等, 图象数据中的颜色、纹理、轮廓和对象等, 音频数据中的音调、音色和旋律等。要进行基于内容的视频检索, 首先要建立一个合理的视频数据模型, 并在此模型的基础上对视频数据进行分割和特征提取, 然后对视频数据或镜头进行检索。基于内容的视频检索技术具有广泛的应用领域, 如远程监控、多媒体会议、虚拟现实、电视新闻、卫星云图、信息娱乐、交互购物、电影和电视的后期制作等, 因此它将对多媒体业产生很大的冲击。

本文介绍一种基于内容检索的视频数据库系统, 通过对视频数据的建模和分析, 讨论了视频数据分割、代表帧及特征提取, 并提出了一种基于内容的视频检索和浏览方法。

2 视频数据模型及系统设计

数据模型是数据库的描述机制, 它从不同的角度和级别描述数据库的内容以及数据间的联系方式——数据库结构和信息组织方式。其发展已经历了网状模型、层次模型、关系模型和面向对象模型, 而进一步的发展趋势将是多媒体数据模型。

视频数据模型是视频数据直接面对应用的表述, 它也是设计视频数据库的关键环节。通常视频数据可以用帧、镜头和场景来描述。帧是一幅静态的图象, 它是组成

视频数据的最小单位。镜头是由连续记录的一个或一系列帧组成, 它是视频数据的基本结构单元, 镜头常用于表示时间和空间连续的一个动作。场景是由时间和空间相关的一组镜头构成, 它描述的是一个完整的事件。

为了视频检索的需要, 可将视频数据分割成一个个镜头。所谓镜头是指一个有意义的场景及其描述数据, 它包括标识符、区间、属性及其相应值集合。每个镜头都有唯一的标识符, 每一区间可以由起始帧和结束帧来表示, 它代表一个连续的视频帧序列。每个镜头均对应于一个视频帧序列, 其内容通过一组属性及其相应值来表示。因此镜头可定义为:(标识符, 区间, 属性/值或特征值集合)。视频数据库系统是用于管理视频数据并为用户提供面向内容查询的视频软件系统。通常基于内容检索的视频数据库由两个子系统组成, 即数据库生成子系统和数据库查询子系统。每个子系统均由相应的功能模块组成, 如图1所示。

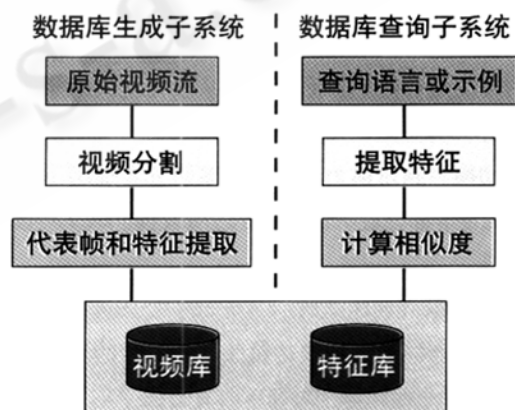


图1 视频数据库的系统结构

数据库生成子系统包括原始视频流、视频分割、代表帧和特征提取等模块。生成的数据库由视频库和特征库组成, 视频库主要包含视频数据或镜头, 而特征库包含用户输入的特征和自动提取的内容特征。数据库查询子系统包括查询语言或示例、提取特征、计算相似度等模块。可以

说, 视频分割、相似性或模糊值是基于内容视频检索的有效方法。

3 视频分割与特征提取

在基于内容的视频检索中, 对视频的处理工作主要包括视频分割、代表帧及其特征的提取等, 其中视频分割又是视频处理中最重要的一步。视频分割是将视频数据分割成一个个镜头的过程, 其核心是识别镜头的切换。镜头切换是一个镜头到另一个镜头的转换, 这时需要确定镜头的边界, 并标出各个镜头的起始帧号和结束帧号。镜头切换主要有两种方式, 即突变和渐变。突变是指一个镜头瞬间直接转换到另一个镜头, 它可使画面的情节和动作发生直接的跳跃, 给人一种轻快、利索的感觉。渐变是指一个镜头到另一个镜头的逐渐过渡过程, 没有明显的镜头跳跃, 从视觉上看镜头比较平滑和舒适。渐变包括淡入(fade in)、淡出(fade out)、隐现(dissolve)和旋转(spin)等。

原始的视频流包含压缩和非压缩的视频数据, 前者提取直流分量, 后者利用邻域平均, 都可得到待检测的视频数据。对于一段视频数据内的相邻画面, 多数都有一些变化, 因此它们的各种特征值也会有一定的变化。经过分析和实验, 帧间颜色直方图差值具有这一特征。由于直方图描述了一幅图象颜色的分布情况, 所以通过直方图的相似度可以判别相邻两幅图象的相似情况, 并以此来判定两帧之间是否有镜头切换。一般采用 x^2 直方图匹配来进行帧间相似度的判定, 当相似度大于某一个值时, 就认为有镜头切换。这种方法对于检测场景变化中镜头的突变是有效的, 而对于场景渐变就失效了。这时可以通过双阈值比较方法来检测场景的变化以获得渐变镜头的切换。到目前为止, 有许多研究者提出了各种视频分割方法, 如改进的直方图、边缘变换、DCT变换、运动矢量等。

视频数据分割成镜头后, 对每个镜头可以提取代表帧, 并用代表帧简洁地表达镜头。代表帧的选取方法很多, 本文采用一种计算相对简单的直方图平均值方法来选取代表帧, 它是将镜头中所有帧的统计直方图取平均, 并选择与该平均直方图最接近的那一帧作为代表帧。代表帧是用于描述一个镜头的关键图象, 它反映镜头的主要内容。对代表帧的特征提取与一般静态图象的特征提取是一样的, 包括颜色、纹理、轮廓和对象等。另外, 镜头还有动态特征, 它们也往往是视频检索时用户所能给出的查询条件, 如镜头的运动变化(摇、推、拉、跟踪等)、运动对象的大小变化以及运动轨迹等。

4 视频检索与浏览

基于内容的视频检索是根据视频的语义特征进行检索, 以提取与特征相符或相似的视频数据或镜头。常用的视频查询方法是通过特定的查询语言或通过可视示例方式来完成。用户要查找一个对象时, 可以用查询语言或示例形成一个查询条件。系统把查询条件中描述的特征转化为具体的特征矢量, 或对示例进行特征提取。将查询描述的特征与特征库中的特征按照一定的匹配算法进行相似度计算, 并返回一组满足一定相似度要求的候选结果。对系统返回的查询结果, 用户可以通过浏览来挑选, 直至得到满意的结果。或者从候选结果中选择一个示例, 经过特征调整后, 又形成一个新的查询条件。这样不断重复操作, 直到用户对查询的结果满意为止。

视频浏览是视频数据库的重要组成部分。当用户对所要查找的目标并不十分明确时, 往往需要对视频数据进行快速的浏览以便寻找感兴趣的内容。浏览的目标是如何跳过一些次要内容, 以较少的图象尽可能多地表达出视频数据的主要内容。通常视频浏览可采用分层结构和集束分类等方法, 并提供方便直观的视频浏览界面。

5 结束语

基于内容的视频检索技术是当前多媒体数据库发展的一个重要研究领域。它的一般方法是将视频数据分割成镜头, 并从镜头中选取的代表帧来表示视频数据。然后再进行特征提取、运动分析以及基于内容的检索等。因此视频分割是视频处理的基础, 目前对镜头的识别和表示还不能充分反映视频中更高层次的内容, 对视频分割仍需作进一步的研究。

视频数据库系统无论在理论上还是在应用上还远未成熟, 很多问题都在研究之中。有些问题还依赖于其他学科的进一步发展, 如数据理论、图象处理、计算机视觉和人工智能等。可以设想, 随着这些问题的逐步解决, 视频数据库系统将在信息社会中发挥重要的作用。■

参考文献

- 1 周润汝等. 视频数据库管理系统导论. 北京: 科学出版社, 2000
- 2 李国辉等. 基于内容的多媒体数据查询和检索. 小型微型计算机系统, 1998, 19(4): 1-8
- 3 曹莉华等. 基于内容检索中的视频处理技术研究. 计算机工程与应用, 1998(6): 39-41
- 4 William I Grosky, et al. The Handbook of Multimedia Information Management. Englewood Cliffs, NJ: Prentice Hall, 1997