

文章编号: 1671-8836(2010)06-0711-06

# 基于 MPEG-7 的视频数据组织与管理

何云峰, 于俊清<sup>†</sup>, 唐九飞, 管 涛  
(华中科技大学 计算机科学与技术学院, 湖北 武汉 430074)

**摘 要:** 针对现有视频数据库系统存在的不足, 根据视频数据的特点, 提出并实现了一种基于 MPEG-7 标准的视频数据组织与管理方法. 该方法将视频内容的描述数据按照不同层次的视频内容单元进行组织, 并针对不同的描述数据, 分别生成结构索引、文本索引和特征索引, 以满足基于内容的视频浏览和检索的需求. 实验证明, 通过视频描述数据库对视频数据进行管理, 不仅降低了服务器的存储要求, 而且为基于内容的视频检索奠定了基础, 具有很好的通用性和可扩展性.

**关键词:** 视频; 描述; 组织; 检索

中图分类号: TP 391 文献标识码: A

## 0 引言

随着视频类型的增加和数据量的日益庞大, 如何有效地组织、管理和传输这些数据, 如何有效地按照视频数据的特性去存取和获取这些数据, 使人们在任何时间、任何地方、任何信息装置, 以人们喜欢的任何方式从海量的视频数据中找到自己感兴趣的相关视频片段已成为一种迫切的需求, 由此引发的相关研究热潮催生了基于内容视频检索<sup>[1]</sup>. 由于视频数据是一种内涵丰富、结构复杂的数据类型, 既具有空间属性又有时间属性, 所以对其进行分析和检索非常困难. 在视频数据的组织和管理方面, 目前被广泛使用的视频数据组织和管理方法主要是基于传统数据库技术的视频管理方法<sup>[2~4]</sup>, 视频数据和描述数据集中存储, 对服务器的存储能力和网络传输能力要求很高, 而且需要大量的人力进行手工标注, 无法解决视频数据的分布式海量存储和实时共享. 随着 MPEG-7 标准(multimedia content description interface, 多媒体内容描述接口)的出现, 相继出现了一些基于 MPEG-7 的研究成果<sup>[5~8]</sup>, 但现有的系统和研究成果主要以特定的、结构化程度比较高的视频类型(如: 电视新闻、报告、专题片或某一种球类视频等)作为研究对象, 并针对这些特定的视频类型

提出了一些特定的解决方案, 缺少通用性.

本文针对现有系统存在的主要问题, 根据视频数据的特点, 提出并实现了一种基于 MPEG-7 标准的视频组织与管理方法, 将视频数据的描述信息, 包括元数据、结构数据、特征数据等数据按照视频层次结构进行组织, 存储在视频描述数据库中, 并针对不同类型的数据分别建立索引, 以满足基于内容的视频检索的需要, 为用户提供方便、有效的基于内容的检索. 通过视频描述数据库对网络中的视频数据进行有效地管理, 不仅可以降低服务器的存储要求, 而且可以基于内容的视频浏览和检索服务提供数据支持, 具有很好的通用性和可扩展性.

## 1 视频描述数据库的组织

在基于内容的视频检索中, 用户检索的不再是视频数据本身, 而是视频的内容. 由于视频数据的复杂性, 视频内容的描述数据具有多样性, 即视频内容可以用不同的方式描述, 例如, 描述视频中的一幅图像, 可以用文本描述, 也可以用颜色和形状特征数据描述. MPEG-7 标准提供了标准的多媒体内容描述方案, 但却没有规定视频内容应该采用哪些信息进行描述. 对于基于内容的视频检索而言, 只有使用同类

收稿日期: 2010-04-05 †通信联系人 E-mail: yjqing@hust.edu.cn

基金项目: 国家自然科学基金资助项目(60703049); 武汉市青年科技晨光计划(200850731353); 华中科技大学自主创新基金(M 2009019)资助项目

作者简介: 何云峰, 男, 博士生, 讲师, 现从事基于内容的视频检索研究. E-mail: yfhe@hust.edu.cn

型的特征信息的视频数据才能够进行有效的检索.因此,视频描述数据库一方面要将视频内容的描述数据进行有效地组织,一方面要为同类型用于检索的描述数据建立索引,以满足基于内容视频检索的需要.

视频数据库系统的总体结构如图 1 所示,原始的视频数据存放在网络中,视频的描述数据存放在视频描述数据库中,二者通过 URL 信息建立联系.

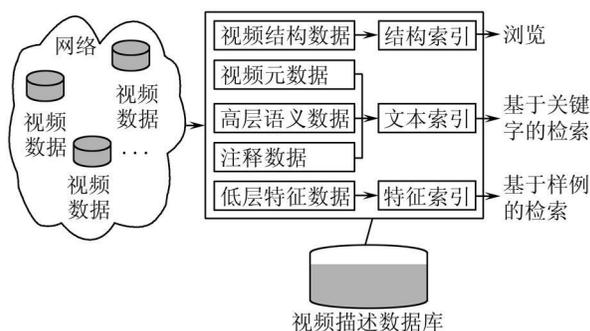


图 1 视频数据库系统的总体结构示意图

### 1.1 视频结构数据

与传统的检索不同,基于内容的视频检索中检索的对象是视频内容单元,如场景、镜头等.为此,需要将视频分割成视频内容单元,由此而产生的数据就是视频的结构数据.按照不同的粒度分割视频,可以得到不同的视频内容单元,它们可以形成如图 2 所示的层次结构,一段视频可以分割为多个情节,一个情节包含若干个场景,每个场景是由一些连续的镜头组成,而镜头包含了视频中连续的多个帧.

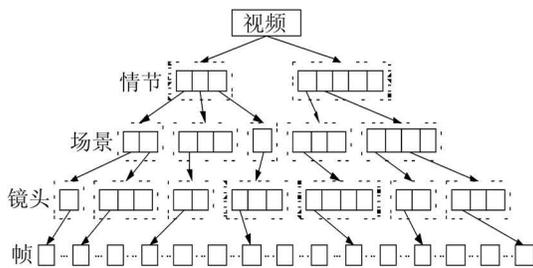


图 2 视频数据的组织结构模型

视频结构信息按照层次进行组织,在视频描述数据库中采用多张关系表进行存储,包括视频表、情节表、场景表、镜头表以及关键帧表等,每张表中包含了父类的 ID、结构 ID、结构类型、起始时间、结束时间以及注释信息等数据项.其中,父类 ID 和结构 ID 用于建立视频结构的层次关系,结构类型则用于描述视频结构的类型,例如镜头的类型(足球视频中的长镜头、中镜头等等)或者场景的类别名称等等.

视频内容单元是视频检索的对象,因此层次化的视频单元组织模型是视频描述数据库组织数据的

基础.在检索的时候,可以通过层次结构不同的视频内容单元进行浏览和检索.

### 1.2 视频元数据

视频元数据主要指视频数据的一些基本特征,它的来源有两个:一类可以通过原始视频数据获取,如视频的格式、视频分辨率信息等;一类是需要由数据库管理员或者是视频数据录入操作者通过手工的方式添加,如视频创建和版本信息、视频标题等等.视频元数据存储在一张关系表中,该表包括元数据名称和元数据值等数据项,并通过视频数据索引值与原始视频建立联系.其中,元数据名称项中存储的是元数据在 MPEG-7 文档中的全路径名称,以最大限度地保留视频元数据的名称信息.此外,元数据值数据项中存储元数据的值,它对应了 MPEG-7 文档中元数据标签中的一个具体值.

### 1.3 高层语义数据

视频的高层语义数据是指运用各种视频处理方法,从视频数据的低层特征以及其他信息,如领域知识等中提取出来的高层语义特征信息,它是视频数据检索中一项重要的文本数据.语义数据在视频描述数据库中采用一个关系表来存储,包括语义信息类型,如行为、人物等,和语义信息值等数据项并通过结构 ID 与具体的视频内容单元建立联系.

### 1.4 注释数据

视频的注释数据是指视频内容的文本描述数据,其来源主要有两种方式:其一是通过手工标注,这种方式的注释数据可以具体到视频内容单元,但具有一定的主观色彩;其二是利用跨媒体信息,即将原始视频数据周围的伴随文本、用户对视频数据的评论等作为视频的注释数据,这种方法可能会产生一定的偏差,即生成的注释数据可能与原始视频并不相关.

注释数据为基于关键字的文本检索方式提供了便利.注释数据根据描述对象的不同,以文本的方式存储在各类视频结构表中的注释信息数据项中.

### 1.5 视频低级特征数据

视频低级特征数据是指用于比较视频内容单元相似度的视频低层特征值,MPEG-7 标准中指定了包括颜色特征、纹理特征、形状特征、运动特征以及声音特征在内的多个低层特征的描述模式,这些特征值大多从属于某个具体的视频内容单元,检索时可以利用特征距离公式计算两组特征值之间的相似度(MPEG-7 标准中提供了大部分的低层特征相似度计算方法).这里需要根据 MPEG-7 标准中提供的特征描述模式,为每一个特征建立关系表,特征值通常是一个向量,向量中的每一个值就是关系表中的一个数



### ① 视频分析处理程序的位置

由于视频数据存在于网络中, 视频分析处理程序可以位于视频数据本地, 也可以在服务器上. 如果分析处理程序位于视频数据本地, 将通过手动调用的方式生成视频数据的描述文档(MPEG-7 文档), 并将该文档上传至视频描述数据库进行后面的入库处理. 如果处理程序位于服务器上的时候, 鉴于目前视频分析处理技术很难达到实时性, 因此在服务器上有一个队列用于存储需要处理的视频数据索引(视频数据在网络上的完整路径名).

### ② 低层特征的提取

低层特征主要是指视觉(包括颜色、纹理和形状)特征、运动特征和音频特征等. 在 MPEG-7 标准的第三和第四部分中, 提供了多类低层特征的提取算法, 可以参考文献[9, 10].

### ③ 结构分析

结构分析是指利用提取出的低层特征以及相关的领域知识, 依据视频数据的组织结构模型, 获得视频数据的层次信息. 结构分析包括镜头检测、场景分割、语义提取等过程.

### ④ 视频表示

视频表示主要是根据 MPEG-7 标准中提供的描述符和描述模式以及描述定义语言对提取出的特征数据和结构数据进行描述并生成 MPEG-7 文档. 由于在 MPEG-7 标准中对于视频内容的单元都采用段(segment), 没有明确给出情节、场景以及镜头的描述模式, 因此需要利用描述定义语言定义这些内容单元的描述模式. 例如, 场景的描述模式可以定义如下:

```
SceneDS
<complexType name="SceneDS">
<complexContent><extension base="mpeg7:DSType">
<sequence>
<element name="MetaDescriptor" minOccurs="0"/>
<complexType>
<complexContent><extension base="mpeg7:DSType">
<sequence>
<element name="MediaTime" type="mpeg7:MediaTimeType"/>
<element name="TextAnnotation" type="mpeg7:TextAnnotationType" minOccurs="0" maxOccurs="unbounded"/>
</sequence>
</extension></complexContent>
</complexType>
```

```
</element>
<element name="Shot" type="ShotDS" minOccurs="0" maxOccurs="unbounded"/>
<element name="Event" type="mpeg7:EventType" minOccurs="0" maxOccurs="unbounded"/>
<element name="ShotRelation" type="mpeg7:RelationGraphDS" minOccurs="0" maxOccurs="unbounded"/>
<element name="EventRelation" type="mpeg7:RelationGraphDS" minOccurs="0" maxOccurs="unbounded"/>
</sequence>
<attribute name="SceneID" type="ID"/>
</extension></complexContent>
</complexType>
```

## 2.3 MPEG7 标准验证

如果可以找到当前处理视频数据的描述文档, 该文档一般与视频数据存放在同一目录下, 此时根据 MPEG-7 标准中提供的描述符和描述模式的定义, 对 MPEG-7 描述文档进行模式验证, 即确定其是否为 MPEG-7 文档, 如果不是则仍然由视频处理模块生成 MPEG-7 文档. 由于 MPEG-7 标准的模式定义语言来源于 XML 模式, 因此 MPEG-7 文档的模式验证与 XML 文档模式验证基本相同, 可以直接用 XML 文档验证器对 MPEG-7 文档进行验证.

确定视频数据描述文档是 MPEG-7 文档后, 为了使该视频数据能够被浏览和检索, 还需要确定该 MPEG-7 文档中是否包含了足够的信息, 包括视频数据的结构数据和特征数据, 这可以通过 XQuery 查询语言对 MPEG-7 文档中的结构和语义数据进行检索, 如果返回的结果集为空, 则需要跳转至视频数据处理模块提取视频数据的结构和特征数据, 并将这些数据加到原有的 MPEG-7 文档中.

## 2.4 MPEG7 文档解析

在获得了视频数据的 MPEG-7 文档之后, 要将 MPEG-7 文档中的信息存入视频描述数据库. 其过程主要是先利用 XML 文档解析器对 MPEG-7 文档进行解析, 获取文档中的数据. XML 文档解析器可以用现有的软件包, 例如微软的 MSXML. 接着在视频描述数据库的视频文件中用 SQL 语句添加一条记录, 主要记录视频文件的名称、网络地址、注释信息以及关键的视频数据索引值, 这个索引值在文档解析的过程中用全局变量来保存. 然后获取 MPEG-7 文档中每个节点的信息, 此时要根据节点的标签以及 MPEG-7 描述符和描述模式判断节点属于那一类信息, 并采取相应的处理.

## 2.5 索引数据

再将视频数据的描述信息存入视频描述数据库

后, 需要生成索引数据, 以便于进行浏览和检索。

索引数据包括结构索引、文本索引和特征索引。其中文本索引和特征索引将按照其对应的索引结构插入相关的数据, 而结构索引需要从原始视频数据中提取出来, 并存储在视频描述数据库中。通常有两类数据用于视频内容的索引。

① 关键帧数据

关键帧图像包含了镜头最重要的信息, 常用于镜头的索引。关键帧图像可以是视频镜头中的一帧或者是多帧, 也可以是由镜头帧构成的全景图。如果在 MPEG-7 文档解析时没有发现镜头中包含关键帧数据, 则根据镜头的信息将中间帧作为镜头的关键帧, 如果有关键帧数据则读取关键帧信息。然后根据关键帧信息, 从网络中原始视频数据中提取出关键帧图像, 将该图像存储在服务器上, 并在视频描述

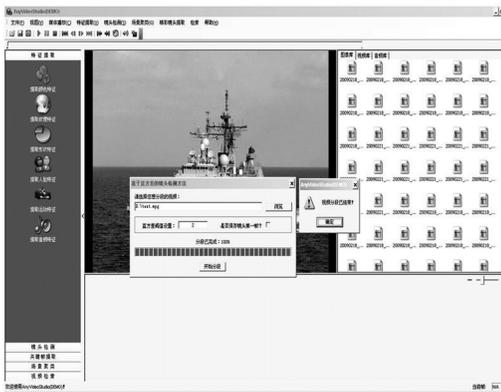
数据库的关键帧表中存储图像在服务器上的路径。

② 视频摘要

视频摘要数据保存了视频数据中重要的内容, 可以作为视频数据的索引。视频摘要的生成主要依赖于视频种类的不同, 例如足球视频和新闻视频有着不同的特点, 视频摘要生成算法也不尽相同。并且视频摘要可以包含多级, 例如文本级别、关键帧构成的图像级别、视频级别等等, 以适应于不同的用户需求。

3 系统实现与分析

按照本文档的设计思想, 基于 VC. NET 集成开发环境设计并实现了如图 5 所示的视频组织与管理



(a) 视频数据的处理



(b) 生成的视频描述文档

图 5 视频数据组织与管理系统示意图

实验选择来自 46 部电影的 1 000 多段视频, 经过视频分段和关键帧提取等视频处理过程, 并对关键帧数据进行了颜色布局( ColorLayout)、边缘直方图( EdgeHist) 特征的提取, 产生的描述数据存储在视频描述数据库中, 其中视频描述数据库采用关系数据库 Oracle 9i。通过实验分析, 本系统主要具有以下特点:

① 通过视频描述数据库对网络中的视频数据进行管理, 降低了服务器的存储要求。

② 通过视频描述数据库组织适合于检索的视频描述数据, 为基于内容的视频检索提供支持。

③ 采用了 MPEG-7 标准用于视频内容的描述, 具有非常好的通用性与可扩展性。

④ 采用视频内容组织模型对视频数据进行组织, 并采用 MPEG-7 标准的模式定义语言定义组织模型中的情节、场景和镜头的描述模式, 使得视频检索可以直接对情节、场景和镜头等视频内容进行检

索。

⑤在视频描述数据库中保存了视频内容数据的索引数据, 包括关键帧和视频摘要, 使得在视频检索中, 可以在用户接口中直接显示索引数据, 而不需要通过网络再对原始视频数据进行处理, 降低了服务器与网络的负担。

4 结 论

视频作为一种典型的非结构化数据, 目前还没有成熟和商用的管理系统, 本文基于 MPEG-7 标准设计并实现了一种较为通用的视频组织和管理系统, 为实现基于内容的视频检索和网络视频搜索奠定了很好的基础。但由于视频的管理涉及内容较多, 算法相对复杂, 其中还有很多细节问题需要解决, 尤其在如何提高分布式网络环境下海量视频数据的检索速度和可靠性方面还有大量的工作需要系统

深入的研究.

## 参考文献:

- [1] Lew M S, Sebe N, Djeraba C, *et al.* Content based multimedia information retrieval State of the art and challenges [ J ]. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2006, 2( 1 ): 1-19.
- [2] Little T D C, Ahange G, Folz R J, *et al.* A digital on-demand video service supporting content based queries [ DB/ OL ]. [ 2009-10-10 ]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.7095&rep=rep1&type=pdf>.
- [3] Weiss R, Duda A, Gifford D K. Content based Access to Algebraic Video [ DB/ OL ]. [ 2009-10-10 ]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.2727&rep=rep1&type=pdf>.
- [4] Oomoto E, Tanaka K. OVID: Design and implementation of a video object database system [ J ]. *IEEE Transaction on Knowledge and Data Engineering*, 1993, 5(4): 629-643.
- [5] Li Honglin, Aahlt S. Hierarchical organization for video annotation [ DB/ OL ]. [ 2009-10-10 ]. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=1627031](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1627031).
- [6] Zhu Xingquan, Elmagarmid A K, Xue Xiangyang, *et al.* Insight video: Toward hierarchical video content organization for efficient browsing, summarization and retrieval [ J ]. *IEEE Transactions on Multimedia*, 2005, 7( 4 ): 648-666.
- [7] Yu Junqing, He Yunfeng, Sunkai, Wang Zhifang, *et al.* Semantic Analysis and Retrieval of Sports Video [ DB/ OL ]. [ 2009-10-10 ]. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=4020977](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4020977).
- [8] Rodriguez A, Guil N, Shotton D M, *et al.* Automatic analysis of the content of cell biological videos and database organization of their metadata descriptors [ J ]. *IEEE Transactions on Multimedia*, 2004, 6( 1 ): 119-128.
- [9] ISO/IEC/JTC1/SC29/WG11 (MPEG). Text of ISO/IEC 15938-3 multimedia content description interface Part 3: Visual. 2002 [ DB/ OL ]. [ 2009-10-10 ]. [http://www.iso.org/iso/catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=40311](http://www.iso.org/iso/catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40311).
- [10] ISO/IEC/JTC1/SC29/WG11 (MPEG). Text of ISO/IEC 15938-4 multimedia content description interface part 4: Audio. 2002 [ DB/ OL ]. [ 2009-10-10 ]. [http://www.iso.org/iso/catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=34231](http://www.iso.org/iso/catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34231).

# Video Data Organization and Management Based on MPEG-7

HE Yunfeng, YU Junqing, TANG Jufei, GUAN Tao

(College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China)

**Abstract:** According to the intrinsic characteristics of video data, a video organization and management method has been proposed and implemented based on MPEG-7 standard, and it has very good versatility and scalability. Video description data are stored in video description database according to difference video content units which are organized by hierarchical mode. And more, video structure indexing, text indexing and feature indexing are constructed for video browse and retrieval. Vast amounts of video data distributed on the network are effectively managed through video description database. It not only reduces the server's storage requirements, but also lays a good foundation for the future Content-Based Video Retrieval.

**Key words:** video; description; organization; retrieval