基于 MPEG-7 协议的视频检索系统设计

薛 玲 李 超 熊 璋

(北京航空航天大学 计算机学院, 北京 100083)

摘 要: 设计了符合多媒体内容描述接口协议 MPEG-7(Multimedia Content Description Interface)标准的电影视频检索系统 SupEye,包括特征提取、数据存储以及特征检索 3 部分. 特征提取负责提取视频特征与摘要,通过分析电影视频特点,给出了视频结构分析及摘要提取的假设、流程及算法. 数据存储负责视频与特征数据的存储,考虑到用户检索历史对检索性能的影响,将 MPEG-7 多媒体特征描述文件分为视频特征描述文件与用户接口描述文件两类. 特征检索构建了一个用于多视频特征检索的 3 层体系结构(表示层、逻辑层和事务层)并给出了检索结果融合流程. SupEye 系统在分析 MPEG-7 标准的基础上解决了视频特征与用户检索特征的分离、样例查询中的特征分析以及多模态查询的结果融合问题.

关 键 词: 基于内容的检索; 视频内容分析; MPEG-7; 特征提取

中图分类号: TP 391

文献标识码: A **文 章 编 号**: 1001-5965(2006)07-0865-04

Design of content-based video retrieval system using MPEG-7

Xue Ling Li Chao Xiong Zhang

(School of Computer Science and Technology, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: A content-based film video retrieval system named SupEye using MPEG-7(multimedia content description interface) was designed, which consisted of three parts: feature extraction, data storage and feature retrieval. The part of feature extraction took charge of the extraction of video features and video abstraction, in which assumptions and algorithms of video structure analysis and video abstraction were described on the basis of analysis on film video features. Data storage focused on the storage of video and feature data, in which multimedia content description files were divided into video features description and user interface description to improve retrieval performance through the usage of users' retrieval features. The part of feature retrieval proposed a three-layered framework for multi-feature video retrieval which was composed of presentation, logic and transaction layer. On the basis of the analysis of MPEG-7, SupEye separates video features from user retrieval features and resolves the problems of feature analysis in sample query and the fusion of multi-feature query.

Key words: content-based retrieval; video content analysis; MPEG-7; feature extraction

鉴于内容组织描述的重要性,动态图象专家组于 2001 年 12 月将 MPEG-7(Multimedia Content Description Interface)定为多媒体内容描述的国际标准协议. MPEG-7 通过定义一系列描述符 D(Descriptor)、描述方案 DS(Description Scheme)与描述语言 DDL(Description Definition Language),提供了一个通用、灵活、可扩展的多媒体内容描述机制 MDS(Multimedia Description Schemes). 其中,D 定义了具体特征属性的描述方式;DS 描述了特征之间的语义与结构关系;DDL 定义了特征的描述语言[1].

1 视频检索系统体系结构

尽管 MPEG-7 已经成为多媒体内容描述的国

收稿日期: 2005-08-18

基金项目: 航空支撑科技基金资助项目

作者简介: 薛 玲(1982-),女,江西南昌人,博士生, scolixue1982@hotmail.com.

际标准,但目前大多数视频检索原型系统仍未兼容 MPEG-7. 鉴于此,本文设计了一种符合MPEG-7特征描述规范的视频检索系统 SupEye,针对电影视频进行检索. 系统体系结构见图 1. ①特征提取子系统. 针对原始的视频数据进行特征提取,形成符合 MPEG-7 规范的特征描述,由视频数据解码、视频摘要自动提取与手动注释 3 部分组成,分别负责原始视频数据的解码、视频摘要的自动提取和视频数据的人工注释. ②数据存储子系统. 管理视频与 MPEG-7 特征的存储与访问控制,由数据存储与数据访问接口两部分组成,分别负责视频数据、视频摘要及 MPEG-7 特征描述的存储和上述信息的读写访问控制及特征相似度匹配. ③特征检索子系统. 解析用户查询要求,对 MPEG-7 特征库进行检索并返回查询结果供用户浏览,由用户接口(portal)、查询条件(parser)分析与查询结果处理 3 部分组成;分别负责需求输入与结果显示,查询实例、查询属性及查询参数的解析和不同查询模式查询结果的融合.

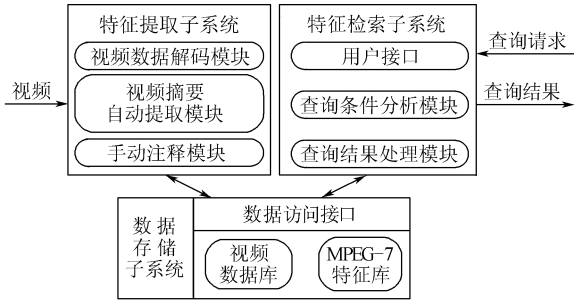

图 1 SupEye 系统体系结构图

2 特征提取子系统

视频摘要自动提取模块是特征提取子系统的重要组成部分,其目标是自动获取视频结构、摘要以及主要的视频特征,工作流程如图 2 所示[2].

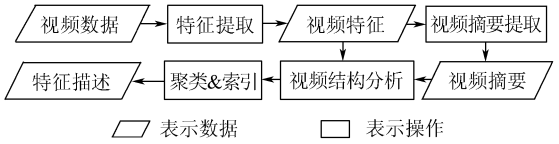

图 2 视频摘要自动提取模块工作流程图

2.1 视频结构分析

视频结构分析的目的是将视频数据划分成语义上独立的场景,便于用户浏览检索结果. SupEye 系统的视频结构分析分两步进行:首先进行镜头边缘检测(shot detection),将视频数据划分成镜头;然后根据镜头之间的相似程度进行场景结构提取.

镜头边缘检测的实质就是检测不同的视频编辑方式. 目前,视频编辑方式有很多,但大致可以划分为 3 类:剪辑(hard cut)、淡入淡出(fades)和溶解(dissolve). 针对上述 3 种不同的编辑方式,SupEye 系统采用 3 种不同的算法进行检测[3]. ①剪辑:颜色直方图算法,剪辑附近的帧差异较大,计算相邻两帧颜色直方图差异检测剪辑;②淡入淡出:色彩饱和度标准方差算法,淡入淡出过程中,象素色彩的饱和度呈线性变化,计算色彩饱和度标准方差的变化趋势检测淡入淡出;③溶解:边缘比较算法,溶解过程中,图像对比度和清晰度达到最低峰值,计算图像中弱边缘相对强边缘的变化检测溶解.

SupEye 系统场景检测的原理是相同场景帧背景的相似性较不同场景帧背景的相似性高. 本文采用文献[4]的方法通过计算主色直方图与空间直方图,检测相邻镜头组之间背景的相似程度来进行场景检测. 此方法不仅考虑了单幅图像的主颜色,而且还考虑了这组图像主颜色的时间与空间变化,抓住了视频作为连续时间媒体的本质.

2.2 视频摘要提取

视频摘要提取步骤在镜头边缘检测之后进行. 其中,代表帧选取是视频摘要提取过程中的关键步骤. 所谓代表帧是指用于描述镜头的关键图像,反映了镜头的主要内容. SupEye 系统的代表帧选取基于以下几个假设进行[5]:①视频中出现的重要的人物应该出现在摘要中;②动作明显的场景在相同的时间内包含了更多的信息,更能吸引人们的注意,因此运动特征明显的帧应该出现在摘要中. 所谓运动特征明显主要体现在两个方面:一方面是指运动强度大,另一方面是指运动集中在帧的中心区域;③带有导演、主演、出品等字样的帧应该出现在摘要中;④电影的主基调确定了电影的类型,反映电影主基调的帧应该出现在摘要中.

针对以上假设,SupEye 系统分别采用了不同的方法进行处理:①采用人脸检测方法,检测特写帧,将其作为代表帧;②采用文献[6]中的基于运动强度描述符与运动空间分布属性的方法,确定视频中运动特征的显著性,将运动特征显著的帧作为代表帧;③通过对导演、出品等字样的检测来获得包含相关敏感信息的代表帧;④电影的主基调主要反映在画面的着色上,将镜头中颜色特征与平均颜色直方图最接近的帧作为代表帧,使提取的视频摘要在画面上更加流畅. 摘要提取流程如图 3 所示.

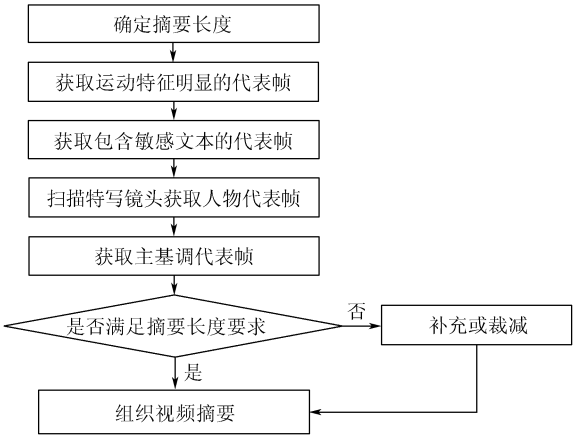

图 3　SupEye 摘要提取流程图

3　SupEye 系统的特征描述方式

为了提高检索的性能,系统设计时充分考虑到用户检索历史对检索结果的影响,因此,在 MPEG-7 特征库中同时记录视频特征与用户检索特征.由于视频特征与用户检索特征围绕的中心、产生的方式不同,SupEye 系统将特征描述文件分为两类:视频特征描述文件与用户接口描述文件,分别描述视频的具体特征和用户喜好及浏览历史,从而实现了视频特征与用户检索特征的分离.具体描述内容如表 1 所示.在对视频进行特征提取操作后,SupEye 形成一个完整的视频特征描述文件,组织结构见如表 2 所示[7-8].

表 1　SupEye 特征描述文件内容

文件类型	描述内容
视频特征描述文件	非视频内容相关:视频标识符、视频格式、视频编码、视频实例等 视频内容相关:视频创作信息、视频段划分、视频段关系、视频摘要、视频基调、视频文本、视频评论
用户接口描述文件	用户标识符、用户喜好、浏览历史

表 2　SupEye 视频特征描述文件组织结构

视频特征描述文件							
视频信息 <Media Information>	视频创作元信息 <Creation MediaInformation>	视频段 <Video Segment>	视频段关系 <Segment RelationshipGraph>	视频摘要 <Summerization>	视频基调 <Affective>	视频文本 <Video Text>	视频评论 <Media Review>
视频标识符 <Media Identification> 视频特性 <MediaProfile>	创作信息 <Creation> 分类信息 <Classification>	时间 <MediaTime> 段划分 <Segment Decomposition> 视频段 <VideoSegment> ⋮ 视频段 <VideoSegment>	段关系描述 <Segment RelationshipNode> 段节点 1 <SegmentNode> 段节点 2 <SegmentNode>	层次型摘要 <Hierarchical Summary> 所属视频段 <ReferenceToSegment> 摘要层次 <HighlightLevel>	情绪类型 <Type> 情绪值 <Value> 文本描述 <Info>	时间 <MediaTime> 文本 <Text>	评论人 <Reviewer> 标准 <RatingCriterion> 评分 <RatingValue> 评语 <FreetextReview>
视频格式 <MediaFormat> 视频编码 <MediaCoding> 视频实例 <MediaInstance>				摘要层次 <HighlightLevel> or 精彩片断 <HighlightSegment>			

根据 XML(Extensible Markup Language)数据模型,表 2 从上至下依次为父/子节点.视频特征描述文件节点(第 0 级)由视频信息、视频创作元信息、视频段、视频段关系、视频摘要、视频基调、视频文本以及视频评论 8 个子节点构成;视频信息节点(第 1 级)由视频标识符和视频特性节点构成;视频特性节点(第 2 级)又由视频格式、视频编码和视频实例节点组成(第 3 级),依次类推.同级节点之间默认为与关系;视频摘要的第 4 级节点中摘要层次与精彩片断节点为或关系.在此选取视频摘要节点为例说明 SupEye 系统的视频特征描述文件结构.

```
<Summarization>
  <HierarchicalSummary name="Sum03"
        SummaryTypeList="keyFrames"
        HierarchyType="independent">
    <ReferenceToSegment idref="segment001"/>
    <HighlightLevel name="coarse keyframe"
          level='0'>
      <HighlightLevel name=
          "medium keyframe"
          level='1' fidelity='e1'>
        <HighlightLevel name="fine keyframe"
          level='2' fidelity='e4'>
          <HighlightSegment name=
            "key-frame E">
            <ImageLocator>
              <MediaTime>……</MediaTime>
            </ImageLocator>
          </HighlightSegment>
        </HighLightLevel>
      ……
      </HighlightLevel>
```

```
    </HighlightLevel>
  </HierarchicalSummary>
</Summarization>
```

4 特征检索

为解决基于样例查询的特征分析与多模态查询的结果融合问题,SupEye 特征检索子系统设计为由表示层、逻辑层与事务层构成体系结构,见图4. ①表示层负责查询条件、查询参数的输入与查询结果的显示. SupEye 设计的用户查询条件包括基于关键字与基于样例的查询;查询参数包括查询条件的匹配方式、相似性阈值以及各特征在结果融合中的比重. 结果显示方面支持层次型结构(hierarchical structure). ②逻辑层由查询条件分析模块与查询结果处理模块组成,分别负责查询请求分解与查询结果融合. ③事务层负责与数据查询子系统进行交互完成特征匹配,该层部署了视频特征匹配、文本匹配等多种子查询组件.

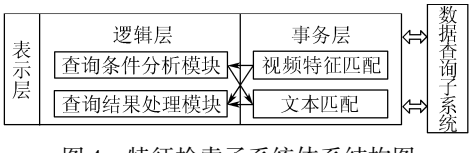

图4 特征检索子系统体系结构图

特征检索子系统工作流程如下:

1) 表示层接收用户查询请求,调用查询条件,将用户输入请求分解为多个子查询.

2) 调用子查询组件,计算各特征向量的距离,将式(1)归一化后获得子查询的相似度值:

$$F_k(q,S_i) = 1 - \frac{D_k(q,S_i) - D_k(q,S_{\min})}{D_k(q,S_{\max}) - D_k(q,S_{\min})} \qquad (1)$$

其中,$F_k(q,S_i)$ 为第 k 种查询组件计算的被查询镜头 q 与镜头 S_i 的相似度值;$D_k(q,S_i)$ 为被查询镜头 q 与视频库中镜头 S_i 的距离;$D_k(q,S_{\min})$ 与 $D_k(q,S_{\max})$ 分别为被查询镜头 q 与视频库中所有镜头距离的最小值和最大值.

3) 将各子查询组件返回结果输入查询结果处理器进行融合,通过式(2)计算总体相似度:

$$F(q,S_i) = \sum_k \lambda_k F_k(q,S_i) \qquad (2)$$

其中,λ_k 为不同查询组件在融合时的权重.

4) 将融合后的最终结果通过用户接口返回用户.

SupEye 系统基于样例的查询界面如图5所示,用户既可选择从样例视频提取何种特征信息用于检索,也可对不同特征在查询结果融合时的比重进行设定[9]. 查询条件分析模块与查询结果处理模块可以根据用户输入调用相应的子查询组件进行特征匹配和结果融合.



图5 SupEye 基于样例查询界面示例

5 结 论

本文在研究现有视频检索原型系统与MPEG-7标准的基础上设计了一种符合 MPEG-7 规范的电影视频检索系统 SupEye,具体内容如下. ①在分析电影视频特点的基础上,详细讨论了视频结构分析与摘要提取的算法及其选取依据;②设计了与 MPEG-7 规范兼容的特征描述文件格式,将多媒体特征描述文件分为视频特征描述文件与用户接口描述文件两类分别用于描述视频特征和用户信息,解决了视频特征与用户检索特征一对多的匹配问题;③针对基于样例查询的特征分析与多模态查询的结果融合问题,设计了一个3层检索系统模型(表示层、逻辑层和事务层).

参考文献 (References)

[1] ISO/IEC JTC1/SC29/WG11N6828, MPEG-7 Overview (version 10)[S]

[2] Dimitrova N, Zhang Hongjiang, Shahraray B, et al. Applications of video-content analysis and retrieval[J]. IEEE MultiMedia, 2002, 9 (3): 42-55

[3] Lienhart R. Comparison of automatic shot boundary detection algorithms[C]// Yeung, Minerva M, Yeo, et al. Proceedings of the SPIE. San Jose CA: Storage and Retrieval for Image and Video Databases VII, 1998: 290-301

[4] 林通,张宏江, 封举富, 等. 镜头内容分析及其在视频检索中的应用[J]. 软件学报,2002,13(8):1577-1585
Lin Tong, Zhang Hongjiang, Feng Jufu, et al. Shot content analysis for video retrieval applications[J]. Journal of Software, 2002, 13(4): 1577-1585 (in Chinese)

[5] Rainer Lienhart, Silvia Pfeiffer, Wolfgang Effelsberg. Video abstracting[J]. Communications of the ACM, 1997, 40(6):55-62

[6] Narasimha R, Savakis A, Rao R M, et al. Key frame extraction using MPEG-7 motion descriptors[C]//Matthews M B. Systems and Computers. California: Pacific Grove, 2003: 1575-1579

[7] ISO/IEC JTC1/SC29/WG11/M6156, MPEG-7 Multimedia Description Schemes WD (Version 3.1)[S]

[8] ISO/IEC JTC1/SC29/WG11/M6156, MPEG-7 Multimedia Description Schemes XM (Version 3.1)[S]

[9] Dunlop M D, McDonald K. Supporting different search strategies in a video query interface[C]// Karen Milligan. Content-Based Multimedia Information Access. Paris: Kluwer Academic Publish, 2000: 21-31

(责任编辑:彭 徽)