

视频摘要技术及其在数字图书馆中的应用

杨 雪

李 馨

孙琳琳

(海南医学院图书馆
海□ 571101)

(吉林大学白求恩医学院病理
生理教研室 长春 130021)

(北京大学法学系
北京 100089)

〔摘要〕 介绍视频摘要的概念及分类,并对现有典型视频摘要系统进行对比分析,着重阐述视频摘要在数字图书馆中的应用,包括应用领域及在大学数字图书馆中的应用,展望视频摘要技术可能的发展趋势和研究方向。

〔关键词〕 视频摘要;数字图书馆;关键帧;视频

Video Abstraction Technology and Its Application in Digital Libraries YANG Xue, Library of Hainan Medical College, Haikou 571101, China; LI Xin, Pathophysiologic Teaching and Research Department, Norman Bethune College of Medicine, Jilin University, Changchun 130021, China; SUN Lin-lin, Law Faculty, Peking University, Beijing 100089, China

〔Abstract〕 The paper elaborates the concept of video abstraction and its classification, comparatively analyzes the current typical systems, focusing on the utilization of this technology in digital libraries, including application areas as well as the application in college and university digital libraries, drawing a blueprint for possible development trend and research directions of video abstraction technology.

〔Keywords〕 Video abstraction; Digital library; Key frame; Video

视频摘要就是通过对视频的结构和内容进行分析,在视频的浏览与检索过程中生成有效的摘要,以便其在网络上进行快速的传输。通过传输视频摘要,用户可以在浏览少量内容的情况下对视频有一个基本的了解。所以视频摘要对于视频传输以及媒体资源管理都有重要的意义^[1]。目前,由于网络上视频资源的数量迅速增加,加工视频结构来有序地提供视频内容浏览和检索服务变得十分重要。其中,基于内容的视频分析技术更是重中之重。如何实现基于内容的视频分析技术,对视频图像进行组织、表达、存储、管理、查询和检索,是数字图书馆面临的一个挑战。

1 视频摘要分类

根据表现形式的不同,视频摘要可分为静态的视频摘要和动态的视频摘要。静态的视频摘要又称为视频概要 (Video Summary),是以静态的方式来表现视频的内容。它是从视频流中抽取或生成的有代表性的图像。动态的视频摘要又称为缩略视频 (Video Skim),是图像序列及其伴音的集合,它本身也是一段视频,但比原视频要短得多^[2-3]。其进一步细分,见图 1。

〔收稿日期〕 2010-07-07

〔作者简介〕 杨雪,助理馆员;李馨,2010 级在读博士。

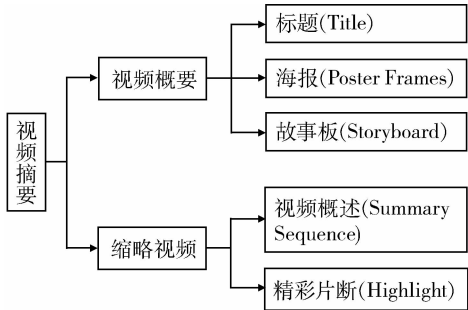


图 1 视频摘要分类

2 典型视频摘要系统及其对比

目前比较典型的视频摘要系统有卡内基梅隆大学的 Informedia 工程、德国曼海姆大学的 MoCA 系统以及美国 FX Palo Alto 实验室研制的 VideoManga 系统^[4-6]。上述视频摘要对比，见表 1。

表 1 典型视频摘要系统对比

系统名称	针对类型	摘要表现形式	重要度规则	特 点
MoCA	影片	HTML 页面	人脸帧、对话帧文本、爆炸、枪响等特殊事件	适合于长影片摘要，可用于网上传输
Informedia	新闻、科教片	缩略视频	人脸帧、字幕帧、运动镜头、场景开始帧	引入语音识别的脚本进行分析生成动态的视/音频缩略视频
VideoManga	会议视频	漫画书	片段长而且新颖	关键帧的大小取决于其重要程度的大小，视觉上符合用户感知

3 视频摘要在数字图书馆中的应用

3.1 视频摘要在数字图书馆的应用领域

3.1.1 视频摘要和浏览 视频摘要是对视频进行浓缩后得到的，保留重要的部分而跳过不重要的部分。故而可实现视频和音频的快速浏览。这样用户在缩短浏览时间的同时仍能得到整个视频片断的完整理解。

3.1.2 视频摘要和检索 视频所包含的内容很复杂，对视频的检索通常要产生大量的检索结果集合，对检索集合的处理方法通常有两个主要途径：（1）尽量返回小的视频片断。以往要从一个视频片断中查找相关的内容，不得不把整个视频片断都浏览一遍。如果能够尽量压缩集中视频片断，就能增加用户的查准率和检索速率。这个经压缩集中的视频片断就是视频摘要。（2）检索结果集的可视化。提取视频的特征，将检索结果集中的对象按照某方面的特征分组并形象化显示，形成结果集，以便在可视化的过程中应用，正是视频摘要所要研究的内容^[7-10]。

3.1.3 视频摘要和索引 由于视频资料内容十分

复杂，用原有文本检索方法检索的查准率会很低。为了更好地满足用户的需求，克服文本检索技术的局限性，需要充分利用视频本身的特征，如颜色、纹理、运动等来建立索引。由于一段视频是由很多帧组成的，对每一个帧都作索引的工作量太大，只能在视频摘要的基础上对原视频建立索引。

3.2 数字视频图书馆系统应用实例

目前最具代表性、技术最为成熟的数字视频图书馆系统是之前提到的卡内基梅隆大学研究和开发的 Informedia 数字视频图书馆。该系统综合了视频分析、图像处理、语音识别、自然语言处理的最新技术，展示了计算机多媒体信息处理的无限空间。

Informedia 系统含有大量视频数据，这些视频材料主要来源于电视台的新闻、科教节目、纪录片，学校的教学录像、课堂录像等录像资料，具体包括 CNN 新闻电视匹兹堡 WQED 台撰稿的广播文献，英国 BBC 为英国公立大学制作的教育文献等。另外还有大量的音频、文本、图像信息，构成了一个庞大的、内容丰富的多媒体信息库^[11]。Informedia 系统分成创建和使用两部分。创建阶段，见图 2。

多媒体信息包括视频、音频、文本等。将多媒

体信息进行数字压缩, 然后进行语音识别、图像处理、自然语言处理等加工, 再进行分割, 存储到数据库中, 并建立索引。创建是一个离线的、后台的过程, 对 1 个小时的原始视频进行压缩、语音识别、索引、分割等都需要花费几个小时的时间。而在该系统的使用方面, 用户的浏览、检索访问等使用操作都是在线、实时进行的, 类似目前的文本数据库使用方式。许多其他的数字图书馆和多媒体数据库都是采用这样一种离线创建、在线使用的工作模式和流程^[12]。

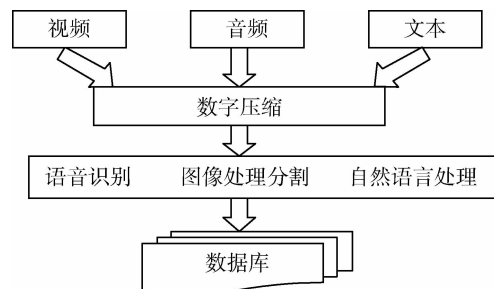


图2 Informedia 系统创建过程

3.3 视频摘要在大学数字图书馆中的应用

由于大学图书馆存储的视频资料主要是用于网络教学的教学材料, 所以它的视频摘要要有自己的特征。这些教学材料主要有两种, 第 1 种是以前的教学录像等视频材料, 第 2 种是配套的教学资料如 PowerPoint 文档等。处理第 1 种材料, 主要是对其进行数字化, 然后运用视频分段、索引等常见的手段进行处理。由于教学视频的特点, 教学视频中教师头像本身并没有携带太多信息, 教学内容主要通过教师的声音实现, 辅助工具本身就是数字化的 PowerPoint, 不必进行特殊处理, 可直接通过手工或半自动的方式建立音频记录和幻灯片文件的同步关系。教学视频资料的处理, 见图 3。

鉴于视频材料的丰富性, 对于每一种视频资料创建视频摘要的方式都会有所不同。(1) 科教片类型的视频材料, 视频分段和关键帧的提取与前面介绍的 Informedia 系统相同。(2) 教师上课型的视频资料, 考虑到大多数镜头对准教师, 并不做切换, 简单根据镜头切分来视频分段不太合适, 容易产生

过大的分段。考虑这些视频材料语意信息大多数保存在语音中, 可以采取一种半自动的方式。根据课程本身的章节性, 预先给每一段定几个关键词, 通过统计脚本中关键词的出现频率, 确定视频材料的分段边界。考虑到仅有教师头像的视频帧含有较少的语义信息, 提取视频摘要和关键帧应该尽量不出现含老师头像的帧。(3) 教学实验型的视频材料, 视频片断可以考虑根据是否出现老师正面头像来分, 因为出现老师正面头像的部分很可能是对实验进行阶段性的总结, 具有标志实验下一阶段开始等特性。相应缩略视频摘要的创建过程中, 其音频部分也应该充分选用含教师头像的帧, 以及画面所对应的音频和关键短语。

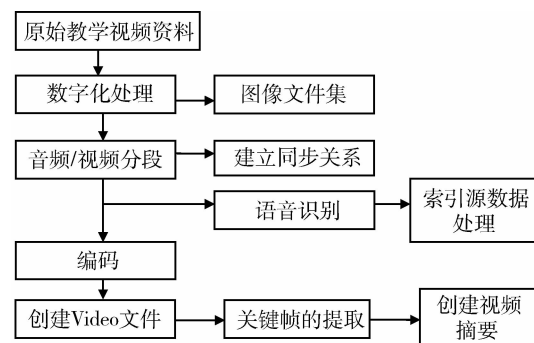


图3 教学视频资料处理

4 结语

远程教学是面向 21 世纪教学改革的重要基础和实质内容, 它能够提供不受地域和时空限制的远程网络教育服务。远程教学是视频技术实现其价值的重要应用领域, 是数字视频图书馆的一个重要应用领域。教师、学生以及有自学自考需要的广大人群将是数字视频图书馆重要的用户群体。另外, 现在的视频搜索引擎发展也比较快。如目前播客搜索引擎新开发出的搜索引擎采用了 BBN 技术公司开发的一种语言系统, 可以将语音转换为可搜索的文本信息, 准确度在 80% 以上。可以把这些搜索引擎应用到数字视频图书馆中, 利用这些搜索引擎的关键技术创建视频摘要以体现音频的主要内容, 总结出音频内容的核心意思, 从而实现用户快速、高查准率的视频音频检索。

(下转第 93 页)