

# 基于内容的流媒体课件检索系统的实现<sup>①</sup>

Realization of Content - Based Retrieve on Streaming Media Courseware

王志梅 (上海交通大学 计算机科学与工程系 200030)  
(温州职业技术学院计算机系 浙江温州 325035)  
周锦成 (温州职业技术学院计算机系 浙江温州 325035)

**摘要:**基于 E-Learning 应用,构建了一套自动和半自动相结合的视频索引和标注机制,使原极其繁琐的索引标注过程变得轻松和简单,并且更加准确;设计了一个基于内容的 PPT 视频课件检索系统的原型,使得用户可以通过知识点和模糊印象迅速地访问到自己感兴趣的视频段。

**关键词:**E-Learning 流媒体 内容检索 PPT 视频

## 1 引言

流媒体课件在 E-Learning 上日益显示出其在表现性、交互性等方面较之 Web 课件资源的优势。在应用中,用户通常希望可以根据自己的具体需求快速找到流媒体课件的某个具体位置(比方对某个概念不甚清楚,希望找到教师正好讲授这个概念的地方)。由于流媒体课件在检索方面比以文本为主的 Web 课件困难得多,因此基于内容的检索和查找(Content Based Index and Retrieval, CBIR)技术在诞生之初就受到各个领域的研究者的关注。本文基于现有的技术,提出了一套适用于 E-Learning 领域的基于内容的检索技术。

本文主要研究对 E-Learning 领域中现存的以 PPT 视频课件为代表的流媒体课件的检索问题。

## 2 基于内容的 PPT 视频课件检索模块组成结构

在大量的视频课件中,学生往往习惯通过输入知识点关键字来寻找某个知识点的视频内容。我们的视频内容检索模块提供了这种功能,学生只要输入知识点关键字进行检索,就能获得若干段与知识点内容相关的独立的视频片段,通过选择相应的片段就能在播放器中播放该片段。学生也可以根据自己对课程的模糊印象,采用多个主题字,对课程内容进行全程查找,

从返回的结果中找到想要的视频片段。

系统的组成如图 1 所示,其中的 Annotation tool 是自动和手动标注模块(将在下面详细介绍)。因为在数据库中查询的速度要远快于直接在视频上的检索,所以本系统将包含内容索引的视频元数据倒入数据库,这样每次检索都会先在数据库进行查询,如果数据库中无法找到,说明视频元数据上并没有建立相应的索引,此时可以将这些关键字纪录在案,使用手工标注的方式完成添加,或者提醒教师下次标注时参考。如果存在相关的内容则将查询结果返回给用户。

整个系统是 B/S 结构的,并将其集成与视频课件点播系统中,以方便学生在听课过程中遇到不清楚的知识点时能及时查询,另外也便于学生按知识点进行学习。图 2 是客户端界面。

## 3 索引及标注模块的实现

索引和标注是实现基于按内容检索的关键步骤。Annotation tool 模块包括以下几部分:

- 抽取教学场景中的高层语义特征:如图 3 所示,我们通过自动视频分割配合 PPT 讲稿来抽取语义特征,并通过人工干预来校正自动抽取的结果。
- 把场景语义和知识点关联:建立一个知识点数据库,然后通过语义的上下文把场景语义和知识点一

<sup>①</sup> 资助项目:科技计划项目(H2005B01)

一对应起来。

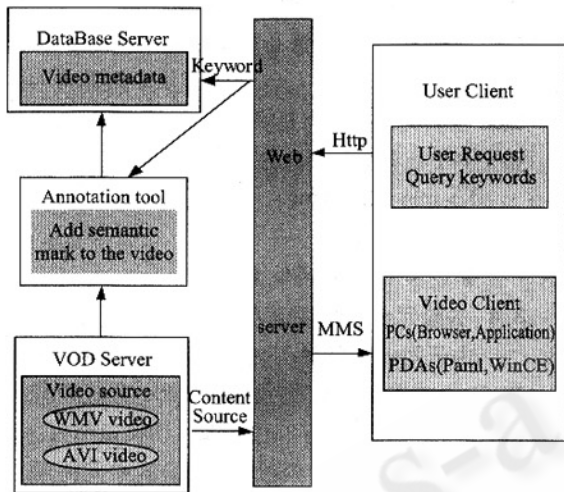


图 1 基于内容的 PPT 视频课件检索模块组成图

构建本系统的关键是如何在视频上构建索引。

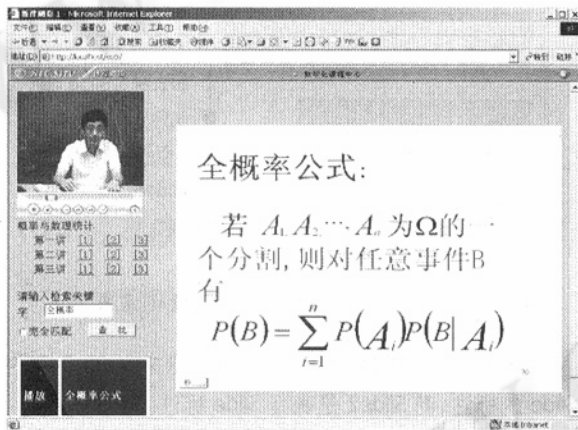


图 2 基于内容的 PPT 视频检索系统客户端界面

· 通过机器学习算法改进搜索结果:通过相关反馈算法来分析用户的行为,调整对应关键字的权值,来优化搜索结果。

如图 3 左边所示,先将实时的视音频流和屏幕教学数据流接入视频分割器,分割成多个镜头(标注的基本单位是镜头和镜头里面的关键帧),然后定义语义词典(语义词典包括 3 个部分:静态场景、关键对象和事件)来描述这些镜头,并且自动抽取 PPT 讲稿中的文本

进行上下文分析作为同步镜头的语义描述。这样在一节课上完后,其初步的视频语义摘要(包括语义描述和特征描述)也就自动生成了;有时候需要手工微调机器自动产生的视频语义摘要或者在后期制作编辑时生成视频语义摘要。如图 3 右边所示,则使用半自动标注工具来人工添加、修改和删除这些标注。

本文,将介绍针对 PPT 视频的另外一种自动标注的方法,即通过 VBA (Visual Basic for Application) for PowerPoint 来完成。

### 3.1 PPT 视频中 Slide 的切换识别

#### 3.1.1 Slide 的切换

由于 PPT 文件的基本组成及浏览单元是 Slide,而且用户在用 PPT 进行演讲时 Slide 的切换频率并不高,因此 PPT 视频结构具有如下的特点:

- 组成每个镜头的所有视频帧都是完全相同的;
- 镜头内没有运动对象;
- Slide 之间的变换多是类似 Cut 之类的突变;
- PPT 中每个 Slide 都具有相对独立的语义内容;
- PPT 视频中的场景往往由一个或多个 Slide 视频段组成;

根据 PPT 视频的结构特点,能否成功的找到每个 Slide 切换时所对应的视频分段(即情节段)的分界,将直接关系到基于内容视频场景构造的效果。因此如何在 PPT 视频中识别 Slide 的切换,即检测情节段的边界是亟待解决的核心问题。考虑到 PPT 视频的特征,解决该问题需要用到全图像切换识别技术。下面首先介绍这种技术,然后阐述用这两种技术完成 Slide 切换检测的设计思想。

#### 3.1.2 Slide 切换识别算法

从 PPT 视频帧中选取用来识别 Slide 切换的代表帧, E-Learning 应用中并不要求检测出 Slide 发生切换的严格边界,即不用把切换边界准确到帧,而只需把切换边界定位在 1-2 秒的时间区域即可。另外一般情况下 PPT 播放过程中 Slide 的切换并不频繁,因此我们按时间每一秒选取一幅代表帧。实践证明,我们使用直方图匹配方法按时间顺序对 PPT 视频进行处理后将识别出所有 Slide 的切换。另外我们针对 PPT 视频硬切换比较多的特点也使用模板匹配的方法,镜头切换的检全率为 100%,但误检率比较高,出错的情况都是 PPT 视频包含动画的情景,阈值需要进一步调整(初始

值为 0.1)。总之,通过镜头检测和视频分割,为接下去按内容的视频场景构造提供了充分的条件。

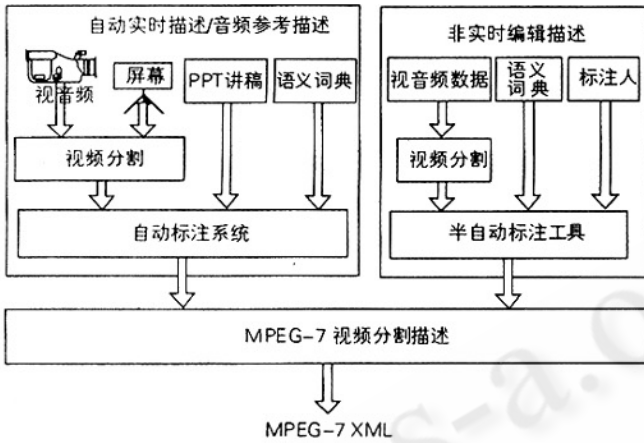


图 3 自动与半自动标注模块框架

### 3.1.3 Slide 的切换的实现

PPT 中每个 Slide 基本上都具有相对独立的语义内容,因此一般情况下每个 Slide 视频段就是 PPT 视频按内容检索的一个基准。

VBA 提供 SlideShowNextSlide 事件来完成 Slide 的切换识别,实现代码略。

### 3.2 Slide 文字抽取

PPT 的每个 Slide 通常有一个主标题,用于对整个 Slide 进行提纲挈领的说明,因此它在内容检索的时候起着非常重要的作用,另外的信息就是其他的用于说明本张 Slide 内容的文字。我们通常也在切换到一张新的 Slide 的同时对文字进行抽取,程序代码略。

按上述方法自动抽取出来的文字将成为本张 Slide 的标注信息,相对应的的时间信息将成为相应 PPT 视频的索引信息。其中主标题通常用于基于关键字的按内容查询。其余文本通常用于模糊查找。因为一张 Slide 通常包含多个文本框,所以保留下来的每个文本信息之间使用分号分开。

### 3.3 去除噪音 Slide 视频标注

在实际的授课过程中,经常会遇到这样的情况:教师跳过几张 Slide 进行讲解,或者往前往后快速翻页。按照上述的方法,这几张 Slide 照样会产生标注和索引。这样会导致学生在进行检索时,虽然定位到了某段 PPT 视频,但是并不符合自己需要。这样的 Slide 称

为“噪音 Slide”,我们需要除去这样的 Slide 标注。

如果将用前述方法产生的所有  $m$  个 Slide 视频标注段按时间排序,则第  $i$  个 Slide 视频标注的持续时间记作  $t_i, i=1, 2, \dots, m$ , 不满足如下限制条件的 Slide 视频标注被认为是噪音 Slide 视频标注,将从整个 PPT 视频标注中除去。

$$t_i > \min\_slide\_time$$

系统中取  $\min\_slide\_time = 5$

### 3.4 去除冗余 Slide 视频标注

由于同一个 Slide 可能会因为噪音 Slide 的出现,造成它在时间维上被噪音 Slide 分成两段出现,对于这种情况,用 2.1 的 Slide 切换识别方法会认为是两段 Slide 视频段,在它们的中间是噪音 Slide 标注。因此在去除所有噪音 Slide 标注后,可能会产生冗余 Slide 标注,即连续出现来自于同一 Slide 的标注,此时应将它们合并成一个完整的 Slide 标注。这种情况如图 4 所示。

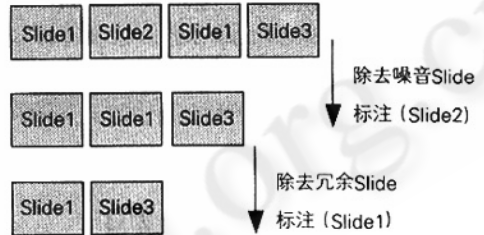


图 4 去除冗余 Slide 视频段示意图

在图 4 中 Slide1, Slide3 分别表示来自于相应的 Slide 的标注,而 Slide2 是噪音 Slide 标注。

检测冗余 Slide 标注非常简单,就是在生成的标注文件中检查是否有两段或者多段连续的标注指向同一个 Slide 的。直接将后面的标注去除即可。

### 3.5 合并纯图像 Slide 标注

纯图像帧是指一个 Slide 没有一行文字,或者只有一个与上一张 Slide 相同的主标题。根据经验,演讲者在使用 PPT 中的纯图像 Slide 进行演讲时,往往是对前面 Slide 内容的补充和详细介绍,因此纯图像 Slide 标注与它的前驱 Slide 标注具有语义内容上的联系,可以将两者合并成一个语义内容上更完整的 Slide 标注。如果某个纯图像 Slide 标注没有前驱 Slide 标注,那它一定

是对后继 Slide 标注的总括或引言,因此可以将它与后继 Slide 标注合并。

按以上规则合并纯图像 Slide 视频段后,也会产生一定数量的冗余 Slide 视频段,需要用 2.4 介绍的方法进行整理。

### 3.6 人工标注

上面所提出的自动标注的方法,并非完美无缺。因为这种方法默认了一个事实:每个教师在制作 PPT 时都是按照一个主标题用于说明一张 Slide 的主要内容,其余文字信息做进一步详细说明。可事实并非如此。另外,本系统通过文字的自动抽取,避免了人工标注仍然需要输入大量文本的问题,而只需要在已有文本的基础上作增加或者删改;通过 Slide 切换的精确定义也解决了人工标注索引建立不够准确的问题。所以本系统保留了人工标注,采用自动标注与人工标注相结合的标注机制。图 5 为本平台提供的修改工具界面截图。

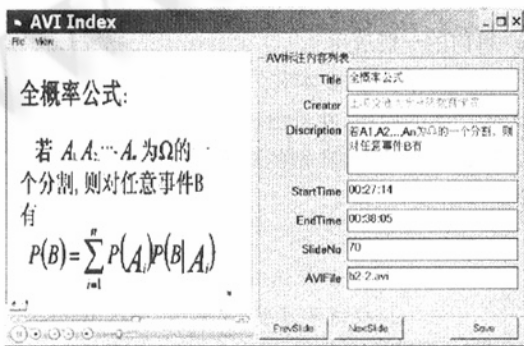


图 5 人工标注程序界面

### 3.7 索引文件

通过以上的标注,系统最终生成的索引文件默认名为 index.xml,其格式如图 6 所示。

## 4 结语

本文所构建的一套自动和半自动相结合的视频索引和标注机制,可以使原有及其繁琐的索引标注过程变得轻松和简单,并且更加准确;同时设计一个基于内容的 PPT 视频课件检索系统的原型,使得用户可以通过文字和图像迅速地访问到自己感兴趣的视频段。

本套系统在网络教育学院的成功应用充分说明,

本文研究得出的算法及系统解决了对 E-Learning 领域中现存的以 PPT 视频课件为代表的流媒体课件检索问题,对于提高学生的学习效率,学习效果具有显著的作用,在应用上具有先进性。

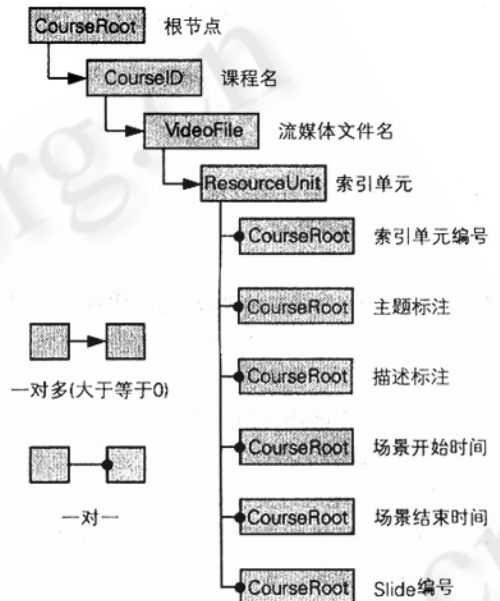


图 6 索引文件格式

### 参考文献

- 1 M. Flickner, H. Sawhney, W. Nibalck, J. Ashley, Q. Huang, B. Dom, M. Gorkhani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The QBIC system. IEEE Computer, 28(9):23-32, Sep 1996.
- 2 M. H. Hayes and M. Jamrozik, "Internet distance learning - The problems, the pitfalls, and the future", Proceedings of IEEE Workshop on Multimedia Signal Processing, pp. 569-574, Copenhagen, Denmark, Sept. 1999.
- 3 <http://www.ctr.columbia.edu/dvmm/>
- 4 <http://vision.ece.ucsb.edu/>
- 5 <http://ltswww.epfl.ch/staff/gu.html>
- 6 章毓晋, 基于内容的视觉信息检索, 科学出版社, 2003。