

基于 MPEG-7 的视频语义检索系统

郑焱 李少波 孙志军 袁婧 王嵩

(中国科学技术大学自动化系 安徽合肥 230027)

摘要 设计和实现一个支持语义的分布式视频检索系统:“语寻”。该系统利用一个改进的视频语义处理工具(该工具基于 IBM VideoAnnEx 标注工具,并增加镜头语义图标注和自然语言处理的功能)对视频进行语义分析和标注,生成包含语义信息的 MPEG-7 描述文件,然后对视频的 MPEG-7 描述文件建立分布式索引,并同时分布式存储视频文件;系统提供丰富的 Web 查询接口,包括关键字语义扩展查询,语义图查询以及自然语句查询,当用户提交语义查询意图后,便能够迅速地检索到感兴趣的视频和片段,并且可以浏览点播;整个系统采用分布式架构,具备良好的可扩展性,并能够支持海量视频信息的索引和检索。

关键词 MPEG-7 语义 视频检索 Lucene 分布式

中图分类号 TP391.4 **文献标识码** A

AN MPEG-7 BASED VIDEO SEMANTIC RETRIEVAL SYSTEM

Zheng Quan Li Shaobo Sun Zhijun Yuan Jing Wang Song

(Department of Automation, University of Science and Technology of China, Hefei 230027, Anhui, China)

Abstract A distributed video retrieval system that supports semantics, called XuNet, is designed and implemented. It uses an improved video semantic processing tool, which is based on IBM VideoAnnEx annotation tool and equipped with shot semantic graph annotation and natural language processing functions to semantically parse and annotate videos and generate MPEG-7 description files that contain semantic information. Afterwards XuNet builds distributed indices for videos' MPEG-7 files, at the meantime distributedly storing video files. The system provides abundant web query interfaces, including keywords semantic expansion query, semantic graph query and natural language query. As soon as a user submits his semantic query intention, XuNet quickly retrieves videos or segments according to his interests or even allows him to browse or play videos on demand. Owing to the distributed architecture, the whole system is highly scalable and supports mass video information indexing and retrieval.

Keywords MPEG-7 Semantic Video retrieval Lucene Distributed system

0 引言

随着计算机处理能力、数字化设备、多媒体、数据库、Web 等技术的不断发展,大量视频信息迅速产生。面对海量的视频信息,如何进行分类、组织、索引以实现海量视频信息的快速检索,就成为人们迫切需要研究解决的问题。目前,传统的视频检索技术大体可以分为两种:一是基于视频的标题、标签、标注等文本描述的元信息的检索;二是基于内容和图像的视频检索^[1]。基于元信息的视频检索技术在互联网上应用得最为广泛,这是由于互联网上的视频分享网站发布的视频一般都包含在网页中,而网页里还包含了有关视频内容的描述信息,这就使得通用的搜索引擎能够利用网页爬虫来获取视频的信息。但是,基于元信息的视频检索技术仅仅依靠人工编辑的文本标注,无法充分获取和利用视频内容本身所包含的信息,因而检索效果有一定局限。基于内容和图像的视频检索技术主要利用计算机图像、视觉等领域的技术,通过对视频内容进行分析处理,将视频分割为场景和镜头两个层次的片段,并提取视频片段中的颜色、纹理、形状、运动等低级特征,最后通过特征的匹配来实现视频的检索。这方面的系统有代表性的是:爱尔兰都柏林城市

大学的 Fischlar 系统^[2]和 IBM 的 Multimedia Search and Retrieval System^[3,4]等。这类技术存在索引建立计算量太大、检索速度慢、检索精确度不高等问题。

目前视频检索领域新的研究热点是基于语义的视频检索技术^[3,4]。基于语义的视频检索是把多媒体信息检索与人工智能技术、自然语言技术相结合的检索。它从语义理解的角度分析和匹配视频对象与检索者的检索请求,是一种基于概念及其相关关系的检索匹配机制。但是如何让计算机自动、高精度地提取视频的语义信息,并且建立海量视频的语义索引,从而达到视频检索语义处理的最终目的,是当前基于语义的视频检索技术研究的热点问题。目前有不少工作关注于从视频低层特征到视频语义描述的自动映射,从而试图解决语义鸿沟问题,这方面有代表性的工作包括:美国 NSF、ASF 和 NASA 资助的数字图书馆项目,IBM 的 CueVideo 项目^[5],卡耐基-梅隆大学的 Informedia Digital Video Library^[6],荷兰阿姆斯特丹大学的 Intelligent Sys-

收稿日期:2011-04-01。科技部支撑计划(2008BAH28B04);安徽高校省级自然科学研究重点项目(KJ2009A152)。郑焱,副教授,主研领域:计算机网络以及网络多媒体。李少波,硕士生。孙志军,硕士生。袁婧,硕士生。王嵩,讲师。

tems Lab 的 MediaMill^[7] 等。但是目前这类利用计算机自动理解和提取视频语义对象及对象间语义关系并建立视频索引的技术仍然不成熟,视频语义对象的识别正确率还很低,远远没有达到商业应用的程度。

目前,在视频机器理解尚未突破的情况下,在商业上最实用的视频索引仍然是建立在视频人工标注基础上的,利用文本索引技术对视频的人工标注建立起索引^[8,9]。而在视频信息的文本描述标准方面,1998年10月,动态图像专家组提出了视频描述接口标准 MPEG-7,该标准可以使用结构化文本(MPEG-7 文件是一种 XML 文件)的形式来描述视频的元信息、低级特征和高级语义特征。与此相配合的,有不少研究机构还提供了支持 MPEG-7 的视频自动分段和标注编辑工具,如 IBM VideoAnnEx Annotation Tool, BitVideo-7 等,这些工具可以自动分析、分段和提取视频的低级特征,并支持人工标注,最后自动将低级特征的描述信息和人工标注的语义信息添加到 MPEG-7 描述文件中。因此,基于 MPEG-7 标准,可以利用工具生成的 MPEG-7 描述文件中的视频语义信息来建立起有效的针对海量视频的语义索引。

然而,如何在 MPEG-7 标准的描述框架的基础上通过人工标注来更好地表征视频语义信息并建立有效的索引呢?为了实现人工标注的视频语义信息的表征,本文提出了利用语义实体和语义实体之间的语义关系来描述视频镜头中的语义信息,其中语义实体采用 MPEG-7 标准中的语义对象(如人物、事件/行为、事物、时间及地点等),语义关系采用 MPEG-7 标准中定义的语义关系(如 agent, agentOf, patient, patientOf, similar, opposite, user, userOf, location, locationOf, time, timeOf 等)^[10,11]。具体来说,将 IBM VideoAnnEx Annotation Tool 改进为一个可视化的语义图标注工具,用户可以利用该工具对视频镜头进行可视化的语义有向图标注(其中语义图的结点为语义实体,有向边为语义关系),然后该工具自动将语义有向图中的所有有向路径进行文本串化生成关键字集,最后写入 MPEG-7 描述文件中。

由于视频镜头的语义图标注人工劳动量太大,本文提出利用 MPEG-7 描述文件中机器标注的或用户标注的视频镜头的自由文本信息进行自然语言分析处理,分析自由文本信息中描述语句的语法结构,提取语义实体及其之间的关系,并同 MPEG-7 标准和知识库中已有的实体信息和关系进行自动匹配,生成语义图描述,最后将文本串化为关键字集并写入 MPEG-7 描述文件中。这样就可以大大地减少人工标注的劳动量,实现自动的语义信息提取。

最后,利用 MPEG-7 描述文件中语义图生成的语义关键字集来建立倒排索引,实现视频的语义索引和检索。

根据以上思路,本文设计和实现了一个支持语义的分布式视频检索系统:“语寻”。该系统提供视频片段级的关键字语义扩展查询、语义图查询以及自然语句查询的功能,并且系统架构可应用支持:(1) 语义镜头表示,通过镜头的语义图人工标注和镜头的自由文本标注的自然语言分析,实现镜头的语义内容的语义图表示,并建立相应的语义文本索引;(2) 镜头级事件查询,用户通过语义图或自然语句来描述“事件”的语义查询意图,在语义镜头表示和索引的基础上实现“事件”的查询。

1 相关技术

1.1 MPEG-7 标准

MPEG-7 全称为“多媒体内容描述接口”,是由动态图像专家组在 1998 年 10 月提出的,目的是要制定一种针对各类多媒体信息的描述标准,这种描述与多媒体内容本身有关,包括低级特征和高级特征^[12]。MPEG-7 描述文件是一种 XML 格式的文件,可以由一些视频处理软件分析生成。

MPEG-7 的主要元素包括以下二种:

1) 描述工具,包括一套标准的描述子 D(Descriptor)和描述方案 DS(Description Schemes)。描述子是指用来定义和表达实体某一方面特征的句法或语法。与视频相关的常用描述子分为颜色(Color)、质地(Texture)、形状(Shape)和动作(Motion)等类型,每一类包含若干种描述子,以不同的方式和结构来描述视觉信息。描述方案是由一个或多个 D 和 DS 构成,DS 规定了它们相互关系的结构和语法。

2) 描述定义语言 DDL(Description Definition Language),用来指定描述方案的一种语言。它是一种模式化语言,是对音视频数据建模结果的一种表征。DDL 规定了 MPEG-7 的描述工具,包括描述符和描述方案,并提供了把描述符构建为描述方案的规则。

在 MPEG-7 描述文件中,视频的描述对象在时间上可以分解为许多镜头,一个镜头分解为多个关键帧。在本系统中,在 MPEG-7 描述文件中的每个镜头加入了三个描述子:自由文本描述子 FreeTextAnnotation 标签(该标签用来写入机器自动标注或用户标注的自由文本信息),语义图描述子 SemanticGraph 标签(该标签用来写入语义图标注经文本串化后的关键字集)以及语义文本描述子 SemanticText 标签(该标签用来写入自由文本经自然语言分析后的关键字集)。

1.2 Lucene

Lucene 是一个开放源程序的全文检索引擎工具包,最初由 Doug Cutting 开发,现在是 Apache 软件基金会项目组的一个子项目。Lucene 使用 Java 开发,是一个完全面向对象的全文检索引擎框架,可以方便地嵌入目标程序中,以实现全文检索功能;同时,开发人员也可以很容易对 Lucene 进行扩展,以定制自己的检索功能。

Lucene 包含文本分析引擎、索引引擎和查询引擎。Lucene 的发行版包含了英文和德文等的分析引擎,其它语言如中文,可以按照 Lucene 提供的文本分析接口,方便地打造各种语言的分析引擎。Lucene 使用的是倒排文件索引方式,其索引文件的格式独立于应用平台。Lucene 的索引可以由一段或多段索引组成,通过索引合并机制,在配置较低的机器上也能高效地索引大量的文件。Lucene 的查询功能非常丰富,不但支持常用的多关键字查询,还支持模糊查询、范围查询、通配符查询等多种查询方式,且自身提供查询解析器,可以处理复杂的查询语法。

1.3 LTP 和语义知识库

哈工大语言技术平台(LTP)是一个语言处理系统框架。它定义了基于 XML 的文本表示,提供了一整套自底向上的语言处理模块,提供了处理结果的可视化工具,并且共享了依存树库、同义词词林扩展版等语料资源。LTP 集成了包括词法、词义、句法、语义等 6 项中文处理核心技术。在 LTP 的基础上,我们能够

跨越语言处理的基本门槛,直接针对高层的课题进行研究。

自然语言处理系统最终需要强大的知识库的支持,因为,自然语言处理首先得理解什么是知识,尤其是关于什么是计算机可处理的知识。知识是一个系统,是一个包含着各种概念与概念之间的关系,以及概念的属性与属性之间的关系的系统^[13]。文献[3]提出应首先建立一种称为知识系统的常识性知识库,它以通用的概念为描述对象,建立并描述这些概念之间的关系,而这些主要是通过知识工程师和专业人员设计编撰而成的。目前,国内重要的知识库包括知网(HowNet)、哈工大同义词林、以及东南大学开发的中文 WordNet。

1.4 基于“知网”的语义相似度计算方法

与传统的语义词典不同,“知网”并不是将每一个概念对应于一个树状概念层次体系中的一个结点,而是通过用一系列的义原,利用某种知识描述语言来描述一个概念,而这些义原通过上下位关系组织成一个树状义原层次体系。因此,可以利用这种知识描述语言表示的两个语义表达式进行相似度计算。

(1) 词语相似度计算

对于两个汉语词语 W_1 和 W_2 ,如果 W_1 有 n 个义项(概念): $S_{11}, S_{12}, \dots, S_{1n}$; W_2 有 m 个义项(概念): $S_{21}, S_{22}, \dots, S_{2m}$ 。 W_1 和 W_2 的相似度是各个概念的相似度之最大值,也就是说:

$$Sim(W_1, W_2) = \max_{i=1, \dots, n; j=1, \dots, m} Sim(S_{1i}, S_{2j}) \quad (1)$$

这样,就把两个词语之间的相似度问题归结到了两个概念之间的相似度问题^[14]。

(2) 义原相似度计算

由于所有的概念都最终归结于用义原来表示,所以义原的相似度计算是概念相似度计算的基础。

由于所有的义原根据上下位关系构成了一个树状的义原层次体系,在义原树中,这条路径的长度就可以作为两个义原的实际距离,所以采用简单的通过语义距离计算相似度的办法。两个义原 P_1, P_2 在这个层次体系中的路径距离表示为 $Dis(P_1, P_2)$ 。计算分两种情况:

① 两个义原处于同一棵树中,路径长度就等于义原距离,语义相似度用下面的公式表示:

$$Sim(P_1, P_2) = \frac{\alpha}{Dis(P_1, P_2) + \alpha} \quad (2)$$

其中 P_1 和 P_2 表示两个义原, $Dis(P_1, P_2)$ 是 P_1 和 P_2 在义原层次体系中的路径长度,是一个正整数; α 是一个可调节的参数^[14]。

② 两个义原处于不同义原树中,可采用的处理方法是义元距离 $Dis(P_1, P_2)$ 一律当作无穷大的常数,即语义相似度为零。

2 “语寻”系统描述

“语寻”是一个基于 MPEG-7 视频描述信息的视频语义检索系统,它通过分析视频的 MPEG-7 描述信息来建立索引,并且既支持通过关键字的语义扩展的方式来检索视频和片段,也支持通过语义图和自然语言来检索语义相似的视频和片段,因此,能够在大量视频中快速定位到视频内部的相关片段,具有检索速度快、精度高的特点。

2.1 系统架构

“语寻”视频检索系统由五个子系统组成,包括:视频预处理子系统、视频语义处理子系统、视频分布式检索子系统、视频

分布式存储子系统、全局信息管理子系统,如图 1 所示。系统提供了三个功能:视频描述信息的分布式索引、视频文件的分布式存储、视频的分布式检索和点播。

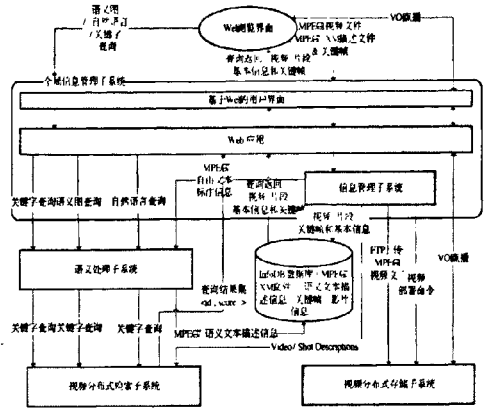


图 1 系统框架图

(1) 视频描述信息的分布式索引

“语寻”将管理员提交的经语义标注的视频 MPEG-7 描述文件和关键帧组成的压缩包进行解压缩,在全局信息管理子系统的控制下对 MPEG-7 描述文件进行 XML 解析,将 MPEG-7 描述文件中有关视频和视频片段的描述信息分别存储到全局信息管理子系统的 InfoDB 数据库中;再将描述信息中的自由文本标注信息提交给语义处理子系统进行处理,得到的视频语义信息再存储到数据库中;最后,将视频和视频片段的所有描述信息传输到分布式检索子系统中进行分布式索引。

(2) 视频文件的分布式存储

视频原文件首先经过视频预处理子系统和视频语义处理子系统的格式转换、切割、语义分析等处理后,由全局信息管理子系统通过 ftp 将处理后的视频文件(.mpg 等格式)上传并最终部署到视频分布式存储子系统中。

(3) 视频的分布式检索和点播

用户通过 Web 界面提交语义查询请求,“语寻”系统首先将语义查询提交给语义处理子系统进行处理,生成包含语义的串化关键字集,接着提交给分布式检索子系统中进行查询,然后将查询结果集返回给 Web 应用服务器,最终显示查询结果给终端用户。用户可以通过视频分布式检索子系统点播视频和视频片段。

2.2 系统用户界面

“语寻”视频检索系统的用户界面主要包括三个部分:管理界面、查询界面、结果呈现和点播界面。

(1) 管理界面

管理界面主要包括:① 提供视频 MPEG-7 描述文件、关键帧以及视频原文件的提交接口;② 提供对分布式存储视频原文件、数据库中的视频描述信息、分布式视频信息索引的查看、删除、更新的同步操作接口。

(2) 查询界面

查询界面要求简单直接并易于用户交互。主要包括三种界面:关键字查询、语义图查询和自然语言查询。关键字查询和自然语言查询提供简洁的输入框接口,语义图查询提供可视化的操作界面,如图 2 所示。另外,系统提供对应的高级查询功能。

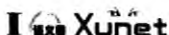



图2 语义图查询界面

(3) 结果呈现和点播界面

系统查询结果界面是由排序的关键帧和描述信息组成的,并且描述信息中关键字高亮显示;用户可以浏览查询结果,预览和点播结果中的视频和视频片段。

3 子系统设计与实现

3.1 视频预处理子系统

视频预处理子系统主要目标是对视频分析处理得到视频内部的描述信息,具体处理过程如图3所示。



图3 预处理子系统处理过程图

视频首先经过预处理模块进行格式转换和切割,得到多个视频片段文件;然后,视频文件片段经特征抽取模块(IBM VideoAnnEx Annotation Tool 视频分析工具),生成 MPEG-7 视频描述文件和关键帧,其中 MPEG-7 视频描述文件内容包括视频的元信息、按照起始时间和长度切割的镜头的低级特征以及一些附加元信息;然后,通过语义标注模块(改进的 IBM VideoAnnEx Annotation Tool 语义图标注工具)对 MPEG-7 视频描述文件中的镜头进行手工和半自动的语义标注,并插入到 MPEG-7 视频描述文件中的镜头中的相应描述子标签中;最后,解析包含语义信息的 MPEG-7 描述文件中的所有信息,并和关键帧一起存储到信息管理子系统的 InfoDB 数据库中。

(1) 手工标注

手工标注包括对视频中切分的镜头进行自由文本标注和语义图标注。自由文本标注是用一段自由文本来描述镜头里面发生的故事,并由标注工具自动插入到 MPEG-7 描述文件中相应的镜头的自由文本描述子标签中。

(2) 语义图标注

语义图标注是语义实体以及实体之间发生的语义关系来描述镜头中的语义信息。在可视化语义图标注过程中,语义实体用结点表示,语义关系用有向箭头来表示,这样,利用一个有向图就可以描述镜头中的语义信息;最后标注工具会自动将语义图中的长度为 0、1、2 级的有向路径串化成字符串关键字集并插入到 MPEG-7 描述文件中相应的镜头的语义图描述子标签中。

(3) 半自动化标注

半自动化的标注主要是利用比较完备的带有时间戳的影片剧本进行标注,将影片剧本中的带有时间戳的场景描述和对白信息按照时间点插入到 MPEG-7 描述文件中相应的镜头的自由

文本描述子标签中,该过程程序自动完成。

3.2 全局信息管理子系统

全局信息管理子系统主要包括 Web 应用服务器、信息管理子系统以及数据库 InfoDB。

(1) Web 应用服务器

Web 应用服务器主要用来接收用户和管理员的请求和返回请求结果,具体功能是:① 将用户的查询和点播请求分别分配到视频分布式检索系统和视频分布式存储子系统中去,最后返回查询结果和流化播放;② 将管理员的视频部署、删除、更新等管理请求提交到信息管理子系统中进行分析和处理,最后返回管理结果。

(2) 信息管理子系统

信息管理子系统主要用来管理整个系统的信息,并处理整个系统的核心业务逻辑,具体功能是:① 对管理员提交的 MPEG-7 描述文件和关键帧的压缩包进行解压和解析,并将解析得到的视频描述信息和关键帧一起存到 InfoDB 中;② 接受管理员的操作请求,对 InfoDB 中的视频和视频片段的信息进行添加、删除、更新等操作;③ 同时,根据②步,发出命令以更新部署在分布式检索子系统的索引信息和分布式存储子系统中的源视频。

(3) 数据库 InfoDB

视频在加入检索系统之前,需要先经过视频预处理子系统和视频语义处理子系统的处理,处理后生成的信息最终包含在一个 MPEG-7 描述文件中以及一组关键帧截图,InfoDB 负责存储这些信息。由于这些信息主要是一些结构化的数据,可以使用关系型数据库来存储(例如 MySQL)。目前,主要信息存储在了两张表中:视频信息表和片段信息表。视频信息表以视频为单位,如表 1 所示。

表 1 视频信息表

列名称	数据类型	说明
video_id	int(11)	影片 ID
title_cn	vchar(255)	中文名称
title_en	vchar(255)	英文名称
publish_time	datetime	发行日期
video_length	int(11)	影片长度
index_status	int(11)	是否被索引
file_size	int(10) unsigned	影片大小
introduction	longtext	简介描述
mdn_status	int(10) unsigned	是否被存储
.....

片段信息表则以片段(镜头)为单位,保存了视频经过分割后产生的每个片段(镜头)的信息,如表 2 所示。

表 2 片段信息表

列名称	数据类型	说明
segment_id	int(11)	视频片段 ID
video_id	int(11)	所属影片 ID
text_annotation	longtext	文本标注信息
keyframe_path	varchar	关键帧存储路径
start	int(10) unsigned	起始偏移位置

列名称	数据类型	说明
length	int(10) unsigned	片段长度
semantic_graph_annotation	longtext	语义图标注信息
semantic_text_annotation	longtext	语义文本信息
.....

3.3 分布式检索子系统

检索子系统是“语寻”视频检索系统的一个分布式的基础服务系统,是在 Lucene 框架基础上进行分布式改造而成的,主要提供视频描述信息的分布式倒排索引和关键字检索功能。系统由 Manager 中心管理节点,Indexer 索引节点,Searcher 查询节点,Collector 归并节点组成,其系统框架如图 4 所示。

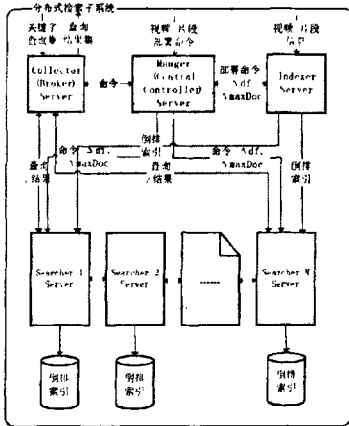


图 4 分布式检索子系统框架图

(1) 视频索引流程

分布式检索子系统在提供用户检索功能之前需要对视频和片段的描述信息进行索引,管理员可以通过全局信息管理子系统的管理界面选择本地要索引的视频的 XML 描述文件和关键帧组成的压缩包,点击“索引”按钮后,系统便开始执行添加索引流程。首先,全局信息管理子系统通告 Manager 索引视频或者视频片段的命令;然后,Indexer 从信息管理子系统数据库中获得 MPEG-7 视频描述信息,建立倒排索引;最后,Manager 按照索引负载均衡通知 Indexer 将索引部署到特定的 Searcher 上。视频索引流程具体如下:

- ① 全局信息管理子系统通知 Manager 一个添加视频索引的任务,其中包含需要添加的 VideoIDs 数组。
- ② Manager 接收到任务后,把 VideoIDs 数组封装成多个任务添加到任务优先队列 TaskQueue 中。
- ③ Manager 上的一个调度线程不断从 TaskQueue 队列中取出一个任务,根据任务类型提交给一个执行线程。
- ④ 执行线程给 Indexer 发送添加视频索引的任务消息。
- ⑤ Indexer 接到通知后,根据 VideoIDs 从全局信息管理子系统中取回对应的视频及其片段的 MPEG-7 视频描述信息,分别建立索引。
- ⑥ Indexer 生成好索引后,通知 Manager 索引完成,并发送 Δdf 和 Δtf 信息给 Manager。
- ⑦ Manager 更新全局 $df = \Delta df, \maxDoc + = \Delta \maxDoc$,并同时通告各个 Searcher 更新它们的全局 df 和全局 \maxDoc ,然后

根据各个 Searcher 上的索引负载均衡计算部署策略,然后将部署策略发送给 Indexer。

⑧ Indexer 根据部署策略,通知指定的 Searcher 将要传送视频索引,通知成功后,开始建立一个连接,传送视频索引。

⑨ Searcher 接收到索引文件后,并将其与本地索引合并完成后,通知 Indexer 索引部署成功。

⑩ Indexer 接收到 Searcher 的成功接收索引的通知后删除本地临时的视频索引,之后通知 Manager 部署成功。

⑪ Manager 上的执行线程在 TaskQueue 队列中添加视频 VideoIDs 完成的任务。

⑫ Manager 上的调度线程从 TaskQueue 队列中取出该任务并通知全局信息管理子系统对应的视频 VideoIDs 已经成功完成部署。

(2) 视频检索流程

“语寻”将关键字查询集请求提交给分布式检索子系统的 Collector,Collector 再向各个 Searcher 节点查询,并将结果归并后返回给 Web 应用服务器,由 Web 应用服务器以合适的方式显示给用户。视频(或片段)查询流程具体如下:

- ① “语寻”的 Web 应用服务器将查询关键字集合成的查询请求发送给 Collector,并根据分页的要求指明返回第 start 条至第 start + count - 1 条搜索结果。
- ② Collector 从 Manager 获取当前在线的所有 Searchers 的 ip 地址,并将查询请求发送给 Searchers。
- ③ Searchers 接收到查询请求,从本地查询,将匹配结果集中的第 1 条至第 start + count - 1 条查询结果的 ids 及其对应的 scores 返回给 Collector。
- ④ Collector 收集 Searchers 返回的结果,对其进行归并、排序,然后将第 start 条至第 start + count - 1 条查询结果 ids 及其对应的 scores 返回给 Web 应用服务器。
- ⑤ Web 应用服务器接收到 Collector 发送来的查询结果 ids,然后向数据库 InfoDB 请求相应的视频(或片段)信息。
- ⑥ InfoDB 接到 Web 应用服务器的请求,将视频(或片段)的关键帧和简介等信息发送给 Web 应用服务器。
- ⑦ Web 应用服务器接收到视频信息后,将其显示给用户。

3.4 语义处理子系统

语义处理子系统由三个模块组成:同义词和语义扩展、语义图字符串化、自然语言处理,并提供两个功能:一是将用户提交的查询进行语义分析和扩展,串化成语义关键字集,再提交给检索子系统检索;二是将信息管理子系统中存储的 MPEG-7 视频描述文件中的自由文本标注信息进行语义分析和扩展,串化成语义关键字集,存储在数据库中,以便随后将该语义关键字集索引到检索子系统中去。子系统框架如图 5 所示。

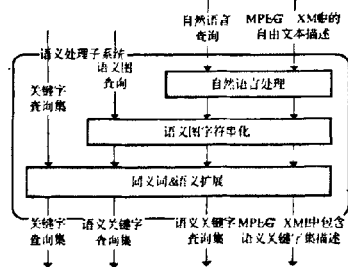


图 5 语义处理子系统框架图

(1) 同义词和语义扩展查询

语义处理子系统同义词和语义扩展模块,主要包括两个功能:一个是查询关键字的同义词扩展;另一个是查询关键字的语义上下位扩展。

系统对提交的每个查询关键字首先利用“哈工大改进版同义词词林”查找到对应的同义词集合,然后利用基于“知网”的相似度计算方法,得到查询关键字和各个同义词的相似度值,该相似度值作为扩展的同义词的查询权重。

接着,系统对每个查询关键字和同义词集合利用中文 WordNet 中定义的分类体系中的上下位关系,进行向上一层的抽象语义扩展和向下一层的具体语义扩展,其中查询关键字和同义词分别扩展的语义词的查询权重值设定不同。

最后,系统再将同义词扩展和语义扩展后的不同查询权重的查询关键字集提交给检索子系统进行检索。

(2) 语义图查询

全局信息管理子系统首先将 InfoDB 数据库中存储的视频片段的语义图标注的语义关键字集传送到检索子系统建立倒排索引。用户在 Web 查询页面上画语义图来表征查询意图,后台 Web 服务器经语义图字符化模块处理为关键字集,并经同义词和上下位扩展,提交给检索子系统进行查询。

(3) 自然语句查询

自然语言处理模块的功能是对现代汉语句子进行一些浅层的语义分析和处理,从而能够在汉语语句中提取出一些语义成分,并以一种规范化的语义图的形式把语义内容表示出来。该模块使用了由哈尔滨工业大学信息检索实验室开发的“语言技术平台 Language Technology Platform (LTP)”工具,以及中文 WordNet 词库。

由于现代汉语句型相对比较复杂,同时相对构造的规则并不能形成一个相对统一的规范,在相关领域对现代汉语句型的研究也比较少。在本模块中,主要只对一部分汉语句型进行分析和处理,处理流程如图 6 所示,处理的句型都是陈述句,具体细分为:“主语+动词”结构、“主语+形容词词”结构、“主语+动词+名词”结构、“主语+动词+名词1+名词2”(双宾语)结构、“主语+动词+名词+动词+名词”(兼语句)结构,以及连动结构等。

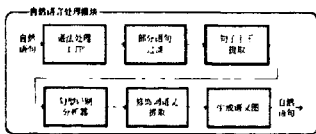


图6 自然语言处理流程

在 MPEG-7 视频描述文件信息存入数据库 InfoDB 时,全局信息管理子系统首先将 MPEG-7 视频描述文件中视频片段的自由文本标注信息传送到自然语言处理模块进行分析处理,得到语义图信息,接着经语义图串化以及同义词和上下位扩展,得到语义关键字集,并存入数据库 InfoDB 中该视频片段的“语义文本信息”字段中,最后语义关键字集传送到检索子系统建立倒排索引。

用户在 Web 查询页面上输入自然语句查询,后台 Web 应用服务器经自然语句处理模块和语义图串化模块处理为语义关键字集,并经同义词和上下位扩展,提交给检索子系统进行查询。

3.5 分布式存储子系统

分布式存储子系统是由媒体分发网络 MDN,以及外围流化万方数据

服务器 MS 集群和 MCP 媒体内容提供商组成,如图 7 所示。媒体分发网络由中心管理模块 MM、目录服务器 DS(多目录服务器),以及若干个分布在网络中的 PN 节点构成的。整个子系统主要功能是提供视频的分布式存储和流化点播功能,具体为:

1) 全局信息管理子系统首先将视频通过 ftp 上传到 MCP 服务器上面,然后通告 MM 部署该视频,MM 根据一定的部署策略(负载均衡和热点调度等)计算并通告相应的各个 PN 到 MCP 上去取对应的视频片段,或者在一定的时机(增加、删除 PN 节点时)按照一定的部署策略进行部署的调整,使得视频内容进入 MDN 的存储区域,以便在 MS 点播用户服务时能够保障服务的质量,并最大限度地提高本地数据的请求命中率。

2) 在 MS 向用户进行流化服务时,如果 MS 发现媒体内容不在本 MS 存储区域,则通过内容分发网络实时请求数据,MDN 系统能够按照一定指标完成媒体内容的实时调度和服务。

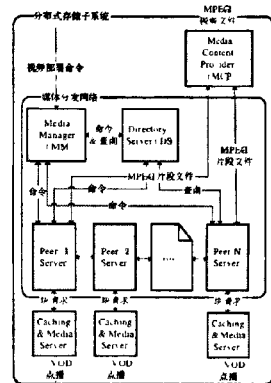


图7 分布式存储子系统框架图

(1) 视频部署流程

系统在提供用户点播功能之前需要对视频源文件进行部署,管理员可以通过全局信息管理子系统的管理界面选择本地要部署的视频源文件,点击“部署”按钮后,系统便开始执行影片部署流程。

① 全局信息管理子系统首先通过 ftp 将视频上传到 MCP 服务器上面。

② 影片上传完成后,全局信息管理子系统给 MM 一个视频部署命令。

③ MM 查询 DS 上面的当前 PN 的存储状况并按照负载均衡和热插均衡等策略计算出视频片段的部署策略,然后通告 DS 和相应的 PN 做好视频片段接收的准备工作。

④ 当 DS 和相应的 PN 上报准备工作就绪之后,MM 通告 DS 创建相应的影片目录管理信息,通告相应的 PN 到指定的 MCP 上去取相应的视频片段。

⑤ 相应的 PN 协同下载相应的视频片段,全部下载完成后,DS 和相应的 PN 上报部署成功的消息。

⑥ MM 接收到所有影片片段的部署成功的消息后,上报给全局信息管理子系统影片部署成功。

(2) 视频点播流程

用户检索到视频片段的列表后,点击感兴趣的视频片段上的“点播”按钮后,浏览器向 Web 应用服务器发出 VOD 请求并刷新页面,该页面自动调用本地的 VLC Media Player,最后播放器主动去连接最近的 MS,开始接收和播放视频流。

① Web 应用服务器的接收到 VOD 请求 (SegmentID, start-

Time, length)后,按照负载均衡将该请求分发到最近的 MS 流媒体服务器上面。

② MS 首先检查本地磁盘是否存在相应的视频片段,如果存在,则直接返回该视频片段所在的位置,否则,MS 向它的 PN 协同下载组中的每个 PN'发出视频块请求(Block Request)。

③ PN'节点接收到块请求后,首先检查本地磁盘是否存在相应的视频片段,如果有一个 PN'存在,则直接多线程将该片段回传给 PN,否则,PN 节点向 DS 查询。

④ DS 查询视频片段的管理目录和数据库信息,并返回给 PN 拥有部分 VOD 请求的视频片段的 PN 结点列表。

⑤ PN 接收到 PN 列表后,向这些列表中的 PN 发出块请求。

⑥ 这些 PN 多线程地将视频片段发送给该 PN,PN 接收到所有片段后,将该片段组装成 VOD 请求所要的视频片段,最后将该片段多线程地发送回给 MS。

⑦ MS 接收到完整的 VOD 请求的片段后,返回给 Web 应用服务器该片段的地址。

⑧ 最后,浏览器上页面内嵌 VLC Media Player 按照指定的视频片段地址连接到 MS 上去流化播放视频片段。

4 结语

本文基于 MPEG-7 标准对视频内容的强大描述能力,设计和实现了一个支持语义的视频检索系统:“语寻”。该系统利用一个改进的 IBM 视频语义处理工具对视频和视频片段进行自由文本的自动和半自动标注、语义图标注、自由文本的自然语句分析,生成包含统一标准的语义关键字集的 MPEG-7 视频描述文件,然后将视频的 MPEG-7 中所有描述信息建立分布式倒排索引,并同时分布式存储视频文件;在查询时,通过多种方式(包括关键字的同义词和上下位扩展查询、语义图查询以及自然语句查询)对视频和片段进行查询,实现对视频内容的高效检索和点播。由于互联网上视频数据快速增长,“语寻”采用分布式架构,具备良好的可扩展性,支持海量视频信息的索引和检索。

今后的研究工作主要是:如何提高语义检索的精度,针对不同用户的个性化视频检索,以及考虑利用用户的评价反馈来改善检索结果的准确度,同时提高系统的容错能力和稳定性。

参考文献

- [1] Baeza-Yates R, Ribeiro-Neto B. Modern information retrieval [M]. England: Addison-Wesley Harlow, 1999.
- [2] Lee H, Smeaton A. The Fischlar digital video recording, analysis and browsing system [C]//Paris: Proceedings of the RIAO 2000-Content-based Multimedia Information Access, 2000.
- [3] Amir A, Argillander J, et al. IBM Research TRECVID-2005 video retrieval system [M]. Washington DC: TRECVID Workshop, 2005.
- [4] Campbell M, Haubold A. IBM research TRECVID-2006 video retrieval system [C]//TREC Video Retrieval Evaluation Proceedings, 2006.
- [5] Syeda-Mahmood T, Srinivasan S. CueVideo: a system for cross-modal search and browse of video databases [J]. Computer Vision and Pattern Recognition, 2000.
- [6] Christel M G. Carnegie Mellon University Traditional Informedia Digital Video Retrieval System [C]//Amsterdam: International Conference

on Image and Video Retrieval, 2006.

- [7] Worring M, Snoek C, Rooij G M D, et al. The Mediamill Semantic Video Search Engine [C]//Acoustics, Speech and Signal Processing, 2007.
- [8] Xu Changsheng, Wang Jinjun, Lu Hanqing, et al. A Novel Framework for Semantic Annotation and Personalized Retrieval of Sports Video [J]. IEEE Transactions on Multimedia, 2008, 10(3).
- [9] Lu Yijuan, Sebe N, Hytinen R, et al. Personalization in multimedia retrieval: A survey [J]. Multimedia Tools and Applications, 2010, 51(1): 247-277.
- [10] Luo Hangzai, Fan Jianping, Zhou Youjie. Multimedia news exploration and retrieval by integrating keywords, relations and visual features [J]. Multimedia Tools and Applications, 2010, 51(2): 625-648.
- [11] Martinez J M. MPEG-7 Overview (version 10) [S]. ISO/IEC JTC1/SC29/WG11, 2002.
- [12] MediaLab S. MPEG-7 White Paper. Outubro, 2003.
- [13] 董振东, 董强, 郝长伶. 知网的理论发现 [J]. 中文信息学报, 2007, 21(4): 3-9.
- [14] 刘群, 李素建. 基于《知网》的词汇语义相似度计算 [C]//第三届中文词汇语义学研讨会, 2002.


(上接第 52 页)

征的标志物投影识别与数据提取算法,该算法主要包括图像背景的消除、Otsu 算法分割图像、图像的形态学运行和最后网格交叉点的提取。实验表明,本文所提出的算法能够有效地提取出校正靶网格线的交叉点,这将为未来 XR/II 图像失真校正打下了良好的基础。

参考文献

- [1] Joskowicz L, Milgrom C, Simkin A, et al. FRACAS: a system for computer-aided image-guided long bone fracture surgery [J]. Computer Aided Surgery, 1998, 3(6): 271-288.
- [2] Fahrig R, Moreau M, Holdsworth D W. Three-dimensional computed tomographic reconstruction using a C-arm mounted XR/II: correction of image intensifier distortion [J]. Medical Physics, 1997, 24(7): 1097-1106.
- [3] Liu R R, Rudin S, Bednarek D R. Super-global distortion correction for a rotational C-arm X-ray image intensifier [J]. Medical Physics, 1999, 26(9): 1802-1810.
- [4] Fantozzi S, Cappello Angelo, Leardini A. A global method based on thin-plate splines for correction of geometric distortion: An application to fluoroscopic images [J]. Med. Phys., 2002, 30(2): 124-131.
- [5] Reimann D, Flynn M J. Automated distortion correction of x-ray image intensifier images [C]//IEEE Nuclear Science Symposium and Medical Imaging Conference Record, Orlando, Florida, 1992: 1339-1341.
- [6] Chakraborty D P. Image intensifier distortion correction [J]. Med. Phys., 1987, 14: 249-252.
- [7] Hoffmann K R, Chen Y, Chen S Y, et al. Pincushion Correction Techniques and Their Effects on Calculated 3D Positions and Imaging Geometries [C]//SPIE, Medical Imaging 96. 1996, 2710: 462-467.
- [8] Sund T, Eilertsen K. An algorithm for fast adaptive image binarization with applications in radiotherapy imaging [J]. IEEE Transactions on Medical Imaging, 2003, 22(1): 22-28.
- [9] Chung K L, Huang H L, Chen I C. New two-phase spatial data structures with applications to binary images [J]. J. Vis. Commun. Image R., 2003, 14(2): 97-113.

基于MPEG-7的视频语义检索系统

作者: [郑焯](#), [李少波](#), [孙志军](#), [袁婧](#), [王嵩](#), [Zheng Quan](#), [Li Shaobo](#), [Sun Zhijun](#),
[Yuan Jing](#), [Wang Song](#)
作者单位: [中国科学技术大学自动化系](#) 安徽 合肥 230027
刊名: [计算机应用与软件](#) 
英文刊名: [Computer Applications and Software](#)
年, 卷(期): 2012, 29(5)

本文链接: http://d.wanfangdata.com.cn/Periodical_jsjyyyj201205017.aspx