

# 建造分布式快速视频交互解析工具<sup>\*</sup>)

Building a Distributed and Interactive Tool for Fast Video Parsing

王伟强<sup>1</sup> 高文<sup>1,2</sup>

(中科院计算所数字化技术实验室 北京 100080)<sup>1</sup>

(哈尔滨工业大学计算机科学与工程系 哈尔滨 150001)<sup>2</sup>

**Abstract** Video parsing is a significant technique to realize content-based retrieval and fast browsing of video documents. Therefore building an efficient and effective tool is a meaningful task. Based on the ideas of intelligent Agents and distributed computation, the paper proposes a framework to build a distributed and interactive tool for fast video parsing, also addresses its implementation. The framework has many advantages, such as quickness, flexibility, scalability and open architecture.

**Keywords** Video parsing, Intelligent agent, Shot, Key frame, Human-machine interaction

## 1 引言

今天,随着计算机处理器性能的不提高,单位容量存储介质价格的不断降低,数字视频在娱乐、教育及其它多媒体应用中发挥越来越大的作用,如视频点播、远程教学、数字图书馆等。由于数字视频信息数量的快速增长以及数字视频信息本身的非结构化特征,对感兴趣的数字视频信息的高效查找变得越来越困难。基于内容的浏览与检索成为人们为实现对视频这样的多媒体信息有效查找而提出的一种新型检索、浏览方式。

一段视频由大量时间上连续的帧组成,每一帧都采用高级压缩技术进行编码存储,如 MPEG 视频编码标准。由于视频媒体数据本身的巨大数据量以及时间上接近的视频帧间的内容往往存在很大的相关性,对一段视频中的每一帧都作索引以便检索、浏览既无必要又妨碍系统性能,因此基于内容的视频检索、浏览系统需要视频解析工具来建立视频流的结构描述,通过对内容的分析抽取出来代表性的关键帧,从而建立索引及内容的表示方式以提供对视频数据的基于内容浏览、检索。

对于一个视频文档,进行结构化分析的第一步便是要探测镜头的边界。镜头是同一个摄像机录制下来的连续的帧序列,它是构成一段视频内容的基本单元。通过对镜头边界的探测,一段视频便被分割成连续镜头构成的有序镜头集。对于通过镜头切分产生的每一

镜头,可以从中选择一个或多个帧来表示该镜头的重要内容,抽取出来的帧被称作关键帧。目前在视频结构分析方面已提出了许多算法,用于切分镜头<sup>[1~4]</sup>,抽取关键帧<sup>[5~6]</sup>等。但目前的视频结构分析技术,如镜头分割,由于各种复杂的场景变化以及编辑效果,还无法做到 100% 的正确率或查全率,因此在计算机自动分割探测过程之后,通过人机的交互来确认、完善分析结果是一种较好的选择。虽然已存在一些在压缩域上的快速分析算法,但由于视频文档的超大数据量,对它进行分析处理仍然是一项耗时的操作,使得整个过程的人机交互显得很很不友好。所以,建立一个具有良好人机交互方式且快速高效的视频结构解析系统是一个有着实际意义的问题。为此我们提出了一种快速视频结构解析的分布式计算框架。

## 2 系统软件结构

视频媒体的大数据量,如通过卫星播放的 MPEG-2 数字电视节目半个小时便会有约 1.3G 字节的数据量,因此在其上运行的各种算法通常是耗时的。当系统运算时,让用户在那里较长时间静静地等待是一种很不友好的人机界面。因此,我们设计了一种分布式的计算模型,一方面它通过并行计算来缩短对一个视频文档进行结构分析的所用时间,另一方面它通过良好的人机交互方式,使人也成为该系统的重要组成部分,它与系统中的计算机进行协同工作,使用户不再陷入

<sup>\*</sup>)本文的研究工作得到国家自然科学基金重点项目(69789301)、国家 863 计划项目(863-306-ZT03-01-2)和中科院百人计划的资助。王伟强 博士生,主要研究领域:多媒体技术、人工智能;高文 教授,博士生导师,主要研究领域:多媒体数据压缩、图像处理、计算机视觉,多模式接口,人工智能、虚拟现实等。

漫长的等待。

整个系统中含有多个智能体,包括有数据读取智能体 DR,码流帧索引智能体 FI,镜头边界探测智能体 SD,关键帧提取智能体 FE,人机交互智能体 HM,如图 1 所示。人机交互智能体 HM 用于实现系统的人机交互,它一方面根据人对系统的要求与相应的智能体进行通讯来提出相应的用户请求,另一方面将镜头边界探测智能体 SD 或关键帧提取智能体 FE 的当前最新产生的结果及时呈现给用户,通过交互让用户对机器的分析探测结果进行审核、修正,并将最终产生的视频文档结构特征的描述数据存储到相应的硬盘上。在对视频文档进行结构分析时,会涉及大量的磁盘数据读取操作,数据读取智能体 DR 负责根据 SD 或 FE 的需要快速从磁盘视频文档中读取相应的数据提供给 SD 或 FE。由于在对视频流进行结构分析时,经常需要对视频码流的特定帧进行读取、帧全部或部分解码、分析操作,且在 MPEG 码流中相邻帧间的数据间隔长度是不等的,为使 DR 具有快速随机访问特定帧的能力,有必要建立各帧在码流中位置的索引信息。在系统中码流帧索引智能体 FI 为 DR 提供任意帧在视频文档中位置的索引信息。码流帧索引智能体 FI 提供对某一视频文档建立帧索引的服务,从而使系统可快速定位随机访问任意帧。镜头边界探测智能体 SD 提供对某一视频文档快速的镜头边界的探测。关键帧提取智能体 FE 通过对视频内容的分析为系统产生的每个镜头抽取合适的帧。整个系统可分为人机交互层、功能层,数据管理层,在图 1 中人机交互层含有 HM,功能层含有智能体 SD 与 FE,数据管理层包含有 DR 与 FI。

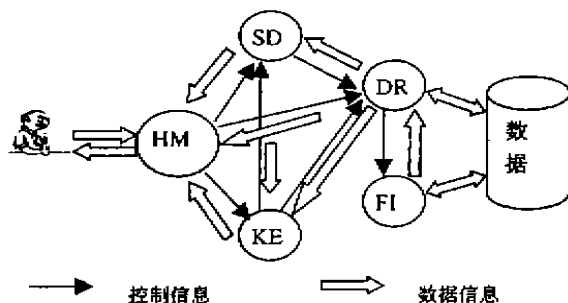


图 1 分布式计算模型的结构

系统中的智能体间主要通过传递消息来进行同步协作。首先,人可以向 HM 下达对某个视频文档进行解析的命令并选择用户工作模式。用户工作模式有两种,离线模式与在线模式。在用户离线模式下,系统对视频文档的分析将处于全自动状态,SD 与 FE 会将机器分析产生的结果存入硬盘中而不再传递给 HM,也

就是说,用户发出解析命令后,便可作其它的事情,以后他随时可通过 HM 从硬盘上察看并调入 SD 或 FE 的以前的输出结果进行审核、修正并将由此产生的视频文档结构描述数据存储到硬盘上。对于在线模式,将根据用户当前选择的处理信息类别由 SD 或 FE 不断将最新产生的分析结果传递给 HM,并由 HM 可视化地表现给用户,由用户作进一步的在线处理,用户与系统中的各智能体并行工作,并不断交换信息。例如,若用户选择的当前处理信息类别为镜头,则解析命令由 HM 向 SD 下达,SD 将不断根据需要请求 DR 读入特定帧的数据,DR 会首先寻找目标文档的索引文件 IndexF,若 IndexF 存在,DR 将利用它定位特定帧在 MPEG 流中的位置,将压缩数据读入缓冲区并传递给 SD,否则 DR 将向 FI 发出请求,FI 将开始为待解析的视频文档建立索引 IndexF,同时在线地为 DR 提供索引信息查询服务。若用户选择的当前处理信息类别为关键帧,则解析命令由 HM 向 FE 下达,FE 将向 SD 索取镜头分割结果,若这时分割结果已存在,则 FE 读取有关镜头分割结果的信息并进行关键帧的抽取计算,否则 SD 将按照上面描述的过程进行镜头边界的探测并不断将新产生的分割结果提交给 FE,同时 FE 对每个镜头都产生相应的关键帧传递给 HM,由 HM 表现给用户,整个过程中的所有相关智能体都在采用类似流水线的方式并行工作并不断交换信息。

### 3 系统实现



图 2 分布式视频结构解析工具的硬件结构

基于上面的思想我们设计实现了一个分布式的快速视频结构分析工具。如图 2 所示,整个计算环境由三台 PIII-450,64M 的 PC 微机组成,它们通过一个集线器构成一个 100M 以太网,各台微机运行的操作系统为 Windows NT 4.0。人们交互智能体 HM 由人机交互工作站来实现,视频结构分析工作站实现了镜头边界探测智能体 SD 与关键帧提取智能体 FE 的功能,数据读取索引工作站实现数据读取智能体 DR 与码流帧索引智能体 FI 的功能。其中码流帧索引智能体 FI 利

用文[10]中的索引模型为某一视频文档建立帧索引,从而使系统可快速定位随机访问任意帧。镜头边界探测智能体 SD 采用我们提出的一种快速镜头边界的探测技术来实现,该技术将另撰文讨论,关键帧提取智能体 FE 采用了文[9]中的自适应关键帧抽取技术。

处于不同工作站的智能体利用 Windows Sockets 来传递消息进行同步协作,处于同一工作站的智能体通过全局变量在线程间传递信息并用事件、信号量、临界区等机制进行同步协作。我们采用如下统一的消息格式(ReceiverID, SenderID, CommandID, DataLength, Data),其中 ReceiverID 为发送智能体的标记, SenderID 为接收智能体的标记, CommandID 为消息类型, DataLength 为消息中 Data 域的字节长度, Data 域的语法及语义随着 ReceiverID, SenderID, CommandID 三者的不同定义而有所不同。例如,当用户向 HM 下达一个对某一视频文档进行镜头分割探测的任务后, HM 便会通过系统启动时 HM 与 SD 间建立的连接性套接字发送镜头分割请求消息(SD\_AGENT, HM\_AGENT, REQ\_SHOTDETECTION, 11, "news0404.mpg"),在目前我们的系统中该消息的数据域为视频文档名。在系统中,若一个消息不携带任何数据作为参数,则 DataLength 域的值为 0 且在消息中 Data 域不存在。

系统可工作在两种模式下,自主模式与受控模式。自主模式下,人机交互工作站将处于临时脱离视频解析系统状态,可从事其它工作。这时处于视频结构分析工作站与数据读取索引工作站中的 SD、FE、DR 及 FI 各智能体将主动地去找寻自己可从事的工作去完成。在我们的系统中,有一个视频文档日志库 VLogLib,库中存储了系统中当前文档状态信息,包括有系统中存在的所有视频文档,已建立的索引文件,自动镜头切分结果文件,自动关键帧抽取结果文件等。各智能体均通过检查 VLogLib 的当前内容来确定自己的当前行为。对于 FI,它将依序检查 VLogLib 中已建立的索引文件信息,一旦发现尚存在某个视频文档,它的索引文件还没有建立,则它便把建立该视频文档的索引作为自己的当前工作,否则若所有文档的索引均建立,则它进入睡眠状态。对于 SD,它将依序检查 VLogLib 中自动镜头切分结果文件的信息,若发现存在某个视频文档,其索引文件已建立,但镜头切分结果文件不存在,则 SD 将为该视频文档建立自动镜头切分结果文件,若找不到符合条件的文档,则 SD 进入睡眠等待状态。FE 与 SD 有类似的选择自己当前行为的方式,只是它的工作条件是某个视频文档的自动镜头切分结果文件已建立,但自动关键帧抽取结果文件还不存在。当 FE 或 SD 处于工作状态时,DR 同时也要处于工作状态,

为它们提供所要求的任意帧压缩视频数据。每当 FE、SD 或 FI 完成了一个任务时,它便要更新 VLogLib 中相应纪录中的特定字段。另外 SD 或 FI 还要向其他的智能体发送 VLogLib 状态更新消息来使那些处于睡眠等待状态的智能体醒来,检测 VLogLib 的状态重新决定自己的行为。当系统处于受控模式下时,将由用户与各智能体协同工作,由人选择进行结构分析的视频文档对象,各智能体将针对该对象根据需要提供自己的服务,随时为用户提供他最需要的数据。

为了提高系统的性能,数据读取智能体 DR 采用预读策略来加速对硬盘数据的读取,即当一个智能体向 DR 发出读取一个大小为 M 的数据块时,当智能体完成对该请求的服务后若处于空闲状态时,则它便会从硬盘中读取在文件中与该数据块顺序相邻的大小为 M 的数据块。

#### 4 系统的特点

本系统将智能体思想与分布式计算相结合,提出了一个快速、灵活的视频解析工具框架,这种框架设计具有如下一些特点。

- 快速。系统采用了一种分布式的结构,将视频结构分析的某一项任务进行了分解,将分解后的子任务映射到不同的物理机上去,使得子任务间可以并发地执行,构成一种高速流水作业模式。例如,系统对某一视频文档执行镜头分割任务时,当视频结构分析工作站上的镜头边界探测智能体 SD 对数据读取索引工作站上的数据读取智能体 DR 发送过来的数据在压缩域上进行解码分析的同时,DR 便从辅助存储器中读取下一个数据块,同时也可能 FI 正在为该视频文件的后续帧建立索引。

- 灵活。系统可存在两种工作模式,若系统处于受控模式,用户当前选择的任务具有优先权,系统中的所有智能体将优先考虑对该任务的服务,这时可能要占用三台机器的资源,用户会得到最佳质量的服务。同时在用户离开系统时,系统有能力自主规划自身的行为,系统一般要占用两台机器的资源。所以,系统具有灵活的适应环境的能力。

- 可扩展性。系统中可轻易加入新的结构分析功能,新加入的功能对应于一个新的智能体,被映射到新的工作站上,可通过扩充现有的消息集实现该智能体与系统中已有的其它智能体间的通讯。即系统的功能层易于加入新的功能项。

- 开放式结构。系统定义了一个清晰的逻辑框架结构,其中组成部件的实现可以基于具有相同功能的不同算法。

**结束语** 视频结构分析是实现视频文档基于内

容浏览与检索涉及的必要的计算过程。由于视频文档的巨大数据量及编码复杂性,使得对它进行结构分析时会存在一个很长的分析计算过程。漫长的等待是一种用户不愿接受的人机交互环境。因此,我们设计了一种分布式的视频结构分析计算环境。它通过多机的并行计算来降低整个过程的时间耗费,利用智能体的思想通过人机的交互协同使得用户获得最佳的服务质量。本系统具有快速、灵活、可扩展性、开放式结构等特点。实践说明我们的设计是正确可行的。

### 参考文献

- 1 Yeo B L, Liu B Rapid Scene Analysis on Compressed Videos. IEEE Transactions on Circuits and Systems for Video Technology, 1995, 5(6): 533~544
- 2 Meng J H, Juan Y J, Chang Shih-Fu Scene Change Detection in a MPEG Compressed Video Sequence. IS&T/SPIE's Symposium on Electronic Imaging, Science & Technology, 1995, 2417(Feb.)
- 3 Kobla V, DeMenthon D, Doermann D. Special effect edit detection using VideoTrails: a comparison with existing

- techniques. In: Proc of SPIE Conference on Storage and Retrieval for Image and Video Databases VII, Jan 1999
- 4 Zabih R, Miller J, Mai K A Feature-Based Algorithm for Detecting and Classifying Production Effects. Multimedia System, 1999, 7: 118~128
- 5 Nagasaka A, Tanaka Y. Automatic video indexing and full-video search for object appearances. Visual Database Systems I, 1992
- 6 Zhang H, Wu J, Zhaong D, Smolian S W An integrated system for content-based video retrieval and browsing. Pattern Recognition, 1997, 30(4): 643~658
- 7 Wolf W. Key frame selection by motion analysis. In: Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc., 1996
- 8 Gresle P O, Huang T S. Gisting of video documents. A key frames selection algorithm using relative activity measure. In: the 2<sup>nd</sup> Int. Conf. on Visual Information Systems, 1997
- 9 Zhuang Y, Rui Y, Huang T S, Mehrotra S Adaptive Key Frame Extraction Using Unsupervised Clustering. In Proc. of IEEE Int. Conf. on Image Processing Oct. 1998: 866~870
- 10 王伟强, 高文. 一种 MPEG-2 流的索引模型及其应用. 已被软件学报录用

(上接第 12 页)

行 What-If 分析预测、资源分配与重组、数据挖掘、企业建模等方向发展。此外,随着网络应用的扩大,支持 Web 平台的应用也是 BI 努力的方向。

### 3. 商务智能在我国的开展

到目前为止,我国对 BI 的研究与开发工作尚处于起步阶段,大多数的研究机构和应用软件开发商仅仅将重点放在对单个应用系统的研究与开发上,缺乏对企业全局运行状况的分析。所开发的应用系统只能对企业运营过程的局部进行管理,缺乏提高企业整体性能的有效手段。针对这种现状,本文对我国开展 BI 研究与开发提出以下几点建议:

1) 尽量将已经成熟的 BI 核心技术引入到现有应用系统中,提高现有系统的技术含量,提高对企业数据利用的有效性,从而提高现有系统的能力;

2) 新系统的研制应注重对企业全局动态性能的分析、评价,不再局限于仅仅作静态数据分析和事务处理;

3) 强调应用系统集成的重要性,使数据在企业组织内外顺利流动。

**结论** 商务智能(BI)作为提高企业整体运行性能的有效手段,正逐步受到企业界和计算机界人士的广泛关注,有着巨大的发展潜力。本文通过对 BI 定义与功能、BI 开发过程与体系结构、BI 发展前景等各个方面的介绍,同时将其与传统决策支持系统进行对比,使大家对 BI 有个全面了解。同时,针对目前 BI 在我国开

展现状,提出了对我国开展 BI 研究与开发的若干建议,希望有助于 BI 在我国的开展。

### 参考文献

- 1 Business Intelligence Strategy--Transforming Your Business Through Business Intelligence. [IBM report]
- 2 Brackett M H. Business Intelligence Value Chain. DM Review, 1999, 3
- 3 BI: What is Business Intelligence?. DM Review, 1999, 3
- 4 The BI Shockwave. Intelligent Enterprise, Nov. 1999
- 5 The Challenge of Business Intelligence Access for the Insurance Industry. SQLiaison Inc, 1999
- 6 Filling the Fact Gap. Next Action Technology, 1997
- 7 White C J. Sybase Adaptive Server IQ: A High-Performance Database for Decision Processing. DataBase Associates International, Inc., January 1999
- 8 The Critical Role of Business Intelligence In E-Business Information Builder
- 9 Teklitz F, McCarthy R L. Analytical Customer Relationship Management. Sybase Inc.
- 10 IBM's Business Intelligence Methodology. Available at: <http://www-3.ibm.com/solutions/businessintelligence>
- 11 Showcase Corporation. Accelerating the Deployment of a Business Intelligence System. 1998
- 12 White C J. The IBM Business Intelligence Software Solution. DataBase Associates International, Inc., Version 3, March 1999
- 13 Symons V. Performance of Business Intelligence Applications using DB2 UDB for AS/400. IBM
- 14 Davis J, Winter R. The IBM Teraplex Integration Centers--Ensuring the Success of Large Scale Business-Intelligence Solutions. Winter Corporation, Waltham, MA, 1999
- 15 Scorecarding for the Enterprise, Cognos
- 16 Maximizing Competitive Advantage with High-end Business Intelligence Technology. OLAP@Work, Inc.