

一种开放式视频管理框架*

刘宇驰¹, 谢毓湘¹, 吴玲达¹, 雷震¹, 戴端辉²

(1. 国防科大信息系统与管理学院, 湖南长沙 410073; 2. 陆军航空兵学院模拟训练中心, 北京 101114)

摘要 对视频进行有效管理的依据是其“内容”。视频具有多模态性,其内容也具有层次性,可用三维模型来描述视频内容。据此,提出了一种开放式视频管理框架。在该框架中,视频中的各层内容与视频数据相对独立地存储与管理,用关联机制建立起视频数据与内容的联系。通过这种机制,能够形成视频数据与内容之间的多种逻辑描述,适用于对多种视频的管理,并且可以对视频进行“多模态查询”。

关键词 三维视频内容模型; 视频管理框架; 视频管理对象; 多模态查询

中图分类号: TP391 文献标识码: A

An Open Framework for Video Management

LIU Yu-chi¹, XIE Yu-xiang¹, WU Ling-da¹, LEI Zhen¹, DAI Duan-hui²

(1. College of Information System and Management, National Univ. of defense Technology, Changsha 410073, China;

2. Center of Simulation Training of Army Aviation Institute, Beijing 101114, China)

Abstract :Video content is the basis for effective video management. As video is characterized by multi-modality and multi-level structure, this paper presents a 3-D model of video content. Furthermore, it proposes an open framework for video management, in which video data and the contents of video are stored and managed separately. The connection between video data and contents was maintained by an associating mechanism, through which multiple logical descriptions between video data and contents could be created. This framework is applicable to the management of different videos while users can search them with “multi-modal query” as well.

Key words 3-D model of video contents; framework for video management; video management object; multi-modal query

对视频数据的有效管理关键在于强大的存储、访问和检索机制。视频往往包含纯视频、音频和文本(字幕),具有多模态性^[1]。另外,视频具有空间和时间维,在组织上的非结构化,数据容量庞大。因此,要将其分解为小的、便于管理的单元,再建立它们之间的联系。本文提出一种开放式的视频管理框架,对视频数据及其内容进行有效的维护、关联与访问。

1 视频管理的研究现状

目前对视频的管理可分为两类:一类是基于时间逻辑结构,另一类则基于语义逻辑结构。

1) 基于时间逻辑结构的视频管理。时间特性是视频管理研究最先考虑的因素,图 1 所示为目前普遍接受的视频的组织结构。一段完整的视频可视为一个故事单元,故事单元是由一系列场景所联系而成的;每个场景又是由若干镜头组成;镜头是由多帧图像顺序构成。镜头可用关键帧来表示,用关键帧的物理特征来建立索引,检索是根据物理特征的匹配来实现的^[2]。镜头又可分成多个子镜头^[3],也可赋予语义^[4,5]。

2) 基于语义逻辑结构的视频管理。其代表是代数视频^[6]。它用节点树来表达视频的层次关系,每个节点表示一段视频的语

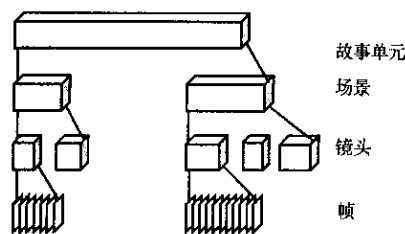


图 1 视频的时间逻辑结构

Fig.1 Video structure on temporal logic

* 收稿日期: 2005-09-08
基金项目: 国家自然科学基金资助项目(60473117)
作者简介: 刘宇驰(1974—),男,博士生。

义概念,由“表达式”来描述。视频代数定义的语义间的逻辑关系突破了镜头的限制,用户可以根据视频的语义内容对视频进行浏览、查找、编辑和表现。另一种管理方式针对视频中的“语义对象”(如人物),首先提取出语义对象,然后建立语义对象、关键帧与原始视频的映射关系,并建立相应的索引树^[7]。

虽然基于语义的方式更加灵活,也更符合人们的习惯,但是很难获取到准确、全面的语义,因此基于时间的方式仍然有很强的生命力。视频管理的趋势是在尽可能地抽取出视频的语义内容的基础上,结合视频的时间逻辑结构,建立起语义、结构与原始视频三者之间的映射关系与管理机制,提供灵活有效的视频访问方法。

2 三维视频内容模型

混合视频的内容可用包含了时间维、媒体维和内容维的三维模型来表述,如图2所示。其中,视频和音频具有低层、中间层和高层内容。在这三类内容的分析基础上,形成混合视频的高层语义内容。

文本包括从视频中字幕探测得到的以及通过语音识别所获得的。音频的低层内容包括音频物理特征,如音量;中间层包括区分出来的语音、音乐和噪音,高层内容则是特定的语义,如某人的讲话。

纯视频的内容也应该是分层的。目前提到的视频低层内容往往指的是单帧图像的物理特征,还应考虑能反映视频动态变化的低层内容,如“全局运动”和“局部运动”。视频的高层内容指的是视频段所表达的含义和概念等语义,具体内容根据不同的应用而定。由于存在着“语义鸿沟”,无法直接把容易获取的物理特征对应于语义,因此在低层与高层内容之间往往要分析出“中间层”内容^[8]。根据不同的应用,也会有相应的中间层内容。

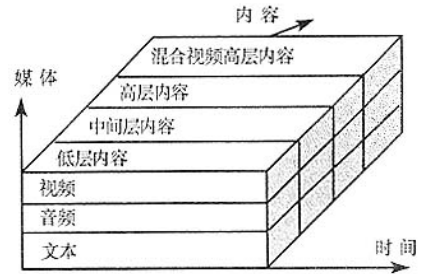


图2 视频内容三维模型
Fig.2 3-D model of video content

3 开放式视频管理框架

在视频管理中,要组织三种数据:视频数据、元数据和内容数据。我们将视频数据及与之关联的元数据和内容数据统称为“视频管理对象”。

3.1 视频管理框架

提出如图3所示的开放式视频管理框架。该框架中,视频管理对象可用三元组表示:

$$VMO = (Vd, Vc, Vm)$$

其中,VMO、Vd、Vc、Vm分别表示视频管理对象、视频数据、视频内容和视频元数据。整个框架由如下构件组成:

- 1) 视频数据管理器负责原始视频数据的管理,用唯一的ID号来标识一个视频文件。
- 2) 视频内容管理器管理视频的各种内容及其与视频的映射关系。
- 3) 元数据管理器管理视频的元数据以及其与视频的映射关系。

视频数据与内容、视频数据与元数据通过相应的映射方法建立起关联。

4) 视图管理器管理的是视图。视图是视频数据与内容之间在语义上或结构上的逻辑关系的表现。有两种视图:①内容视图:与一个视频段所对应的内容的逻辑关系。②视频视图:与内容相关联的视频段的逻辑表示。

5) 查询管理器将用户的查询解析为内容查询模式,对视图和元数据进行查询。

6) 内容字典负责维护视频可能涉及到的内容(特别是语义内容),即术语。

管理员通过视频数据管理器、元数据管理器、视频内容管理器和内容字典来维护相应的数据及视频数据与内容之间、视频数据与元数据之间的联系。

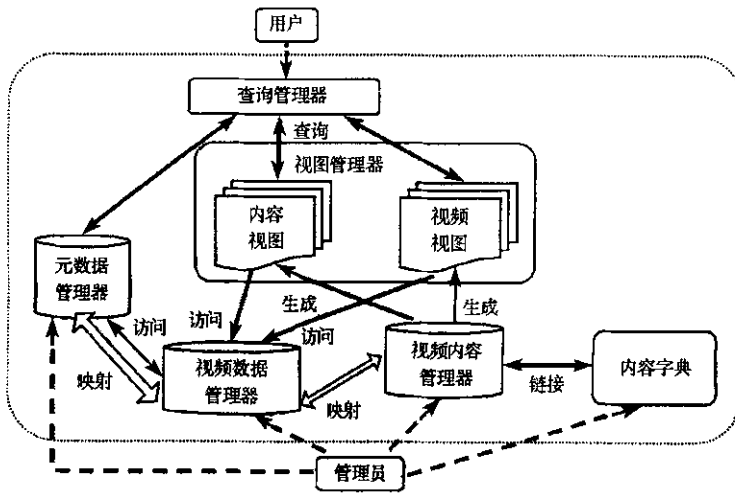


图3 开放式视频管理框架

Fig.3 Open framework for video management

3.2 关联机制

将视频管理对象中的原始数据、内容与元数据分开管理可以保证数据的独立性,不至于对其中一种数据的内部管理而影响到另外两种数据。另一方面,我们还需要我们建立起有效的关联机制,保证视频对象中三种元素的数据的完整性与一致性,并实现协同动作。

1) 映射机制

以视频数据与其内容的映射为例,映射关系可用如下方式建立:

$$\text{Map}(V_c, V_d, T(m, n))$$

$\text{Map}()$ 表示建立映射关系,此式表示将内容 V_c 映射于视频数据 V_d 中起始帧为 m 和结束帧为 n 的一段视频 $T(m, n)$ 。相应的,也要有解除映射的方法。

应用这种映射机制,能够将任意内容映射到任意视频的任意段,实现更灵活、更精细的内容粒度。同样的内容可以映射到一个或多个视频文件中的多个段中,这样可以迅速访问位于不同视频文件中具有相同的内容的视频段。

2) 视图管理

视图是按一定的逻辑概念(如语义逻辑)形成,如下所示:

$$\text{Create}(\text{ContentView}, V_d, \text{SubSet}\{V_c | V_d\})$$

其中, ContentView 表示内容视图, V_d 表示视频数据, $\text{SubSet}\{V_c | V_d\}$ 表示视频数据 V_d 的内容 V_c 的一个子集。这个子集里的内容是根据一定的标准选择的,比如某种语义框架。视频视图可用类似的方法生成。视频内容与视频数据映射关系的变化也会引起视图的变化,这种变化反映在树或图中节点的操作上,即节点的增加、删除、修改。

这种关联机制可以建立视频数据与内容之间的多种逻辑描述,这也使得该框架适用于不同视频的管理,具有开放性。

3.3 多模态查询

理想的查询是基于视频的语义。在语义不完全时,查询还要依靠中层和低层内容。这里,我们提出一种“多模态查询”方法,即根据纯视频、音频和文本的内容进行查询。如:

$$\text{Query}(\text{Shots}, \text{ShotType}(\text{Close-up}), \text{Audio}(\text{Audience-Loud}), \text{Caption}(\text{“Henry”}))$$

它表示查找有近影、有巨大观众声响且有字幕“Henry”的镜头,这可能意味着 Henry 进球了。其具体实现可以在生成的混合了多种模态内容的视图中进行。

4 结束语

本文分析目前视频管理的特点,讨论了视频管理系统中要管理的元素,建立了三维视频内容模型,据此提出一种开放式的视频管理框架,并分析了要研究的关键技术。该框架将视频数据、元数据、内容数据分开存储与维护,保持了数据的独立性,并用关联机制建立起相应的联系,用视图来形成查询的主体。使用这种框架,用户可以根据自己的需求建立符合具体应用领域的视频管理系统。同时由于保留了多层内容,可以在语义不完全的情况下实现“多模态查询”。该框架中许多具体实现技术还有待于我们在今后的工作中进一步深入研究。

参考文献:

- [1] 吴玲达 等. 多媒体信息系统[M]. 北京:电子工业出版社,2002.
- [2] Mahdi W, Ardebilian M, Chen L M. Automatic Video Scene Segmentation Based on Spatial-temporal Clues and Rhythm[J]. Networking and Information Systems Journal, 2000, 2(5):1-25.
- [3] Chen J Y, Taskiran C, Albiol A, et al. ViBe: A Compressed Video Database Structured for Video Active Browsing and Search[J]. IEEE Transactions on Multimedia, 2004, 4(1):103-118.
- [4] Naphade M R, Kozintsev I V, Huang T S. A Factor Graph Framework for Semantic Video Indexing[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(1).
- [5] Zhou W, Vellaikal A, Kuo C C J. Rule-based Video Classification System for Basketball Video Indexing[A]. In ACM Multimedia 2000, Los Angeles, USA, 2000.
- [6] Weiss R, Duda A, Gifford D K. Composition and Search with a Video Algebra[J]. IEEE MultiMedia, 1995, 2(1):12-25.
- [7] Chen L, Ozsu M T, Oria V. Modeling Video Data for Content Based Queries: Extending the DISIMA Image Data Model[A]. In: Proceedings of the 9th International Conference on Multimedia Modeling, 2003:169-189.
- [8] Hsu W, Chang S. A Statistical Framework for Fusing Mid-level Perceptual Features in News Story Segmentation[R]. IEEE International Conference on Multimedia & Expo (ICME) 2003, invited paper.

(上接第50页)

参考文献:

- [1] Proakis J G. 数字通信(第四版)[M]. 张力军 等译. 北京:电子工业出版社,2003.
- [2] Cartwright K V. Blind Phase Recovery in General QAM Communication Systems Using Alternative Higher Order Statistics[J]. IEEE Signal Processing Letters, 1999, 6(12):327-329.
- [3] Serpedin E, Ciblat P, Giannakis G B, et al. Performance Analysis of Blind Carrier Phase Estimators for General QAM Constellations[J]. IEEE Transactions on Signal Processing, 2001, 49(8):1816-1822.
- [4] 韩钢. 自适应单载波、多载波调制中信号盲检测技术研究[D]. 西安:西安电子科技大学,2003,10.
- [5] Gardner W A, Spooner C M. The cumulant theory of cyclostationary time-series, Part I: Foundation[J]. IEEE Transactions On Signal Processing, 1994, 42(12):3387-3408.
- [6] Spooner C M, Gardner W A. The Cumulant Theory of Cyclostationary Time-series, Part II: Development and Applications[J]. IEEE Transactions On Signal Processing, 1994, 42(12):3409-3429.
- [7] Dobre O A, Bar-Ness Y, Su W. Robust QAM Modulation Classification Algorithm Using Cyclic Cumulants[A]. In Proc. IEEE WCNC, Atlanta, GA, 2004:745-748.
- [8] Swami A, Sadler B M. Hierarchical Digital Modulation Classification Using Cumulants[J]. IEEE Transactions on communications, 2000, 48(3):416-429.
- [9] Dandawate V, Giannakis G B. Asymptotic Theory of Mixed Time Averages and Kth-order Cyclic-moment and Cumulant Statistics[J]. IEEE Trans. Information Theory, 1995, 41(1):216-232.
- [10] 张贤达,保铮. 非平稳信号分析与处理[M]. 北京:国防工业出版社,1998.

