视频摘要

陈剑赟 老松扬 吴玲达
(国防科学技术大学多媒体研发中心,长沙 410073)

摘 要 众所周知,文本的摘要是对一篇文章的一个简短的小结.随着对视频数据处理要求的不断提高,出现了类似的概念——视频摘要,同时也产生了视频摘要技术.同文本摘要概念相似,视频摘要是对视频的一个简短的小结.为了使人们对视频摘要有一个概略的了解,该文首先阐述了视频摘要的基本概念,同时简单介绍了视频摘要的应用,并在此基础上对视频摘要进行了分类;然后重点介绍了每一类视频摘要的实现技术——关键帧提取技术、多特征融合技术等;最后对目前视频摘要技术进行了小结,并展望了若干发展途径.

关键词 信息处理技术(510·40) 视频摘要 视频概要 缩略视频 关键帧 镜头 场景
中图法分类号: TP391 TN941.1 **文献标识码**: A **文章编号**: 1006-8961(2003)07-0721-05

Video Abstraction

CHEN Jian-yun, LAO Song-yang, WU Ling-da
(*Multimedia R&D Center of National University of Defence and Technology, Changsha* 410073)

Abstract The abstraction of an article is a short summary of a document. With the development of video processing, it comes into a similar concept—Video Abstraction. The idea of video abstraction is in very much the same way as text: a short summary of the content of a longer video document. This paper first explains the basic concepts of video abstraction, introduces the applications of video abstraction and classifies video abstraction into video summary and video skimming. Then it emphasizes on the realization techniques of video abstraction, including key-frame extraction and multi-features fusion etc. At last this paper makes some conclusions about video abstraction and prospects for some approaches to it.

Keywords Video abstraction, Video summary, Video skimming, Key frame, Shot, Scene

0 引 言

在计算机和通信技术高速发展的今天,随着宽带网络技术、音视频压缩技术以及计算机硬件技术的发展,使得电脑有足够的能力来传输和存储大容量的多媒体数据,并能建立大规模的多媒体数据库.与此同时,对多媒体数据的需求又加速了多媒体数据库技术的发展,而视频媒体因为其信息容量大而在其中处于重要的地位,数字视频技术也得以迅速发展.

视频分析和处理的初期主要集中在分析视频帧的低层特征上,例如颜色、形状、纹理等;而目前的研究则主要集中在更加接近直观内容的分析上[1,2],其中一个重要的研究内容就是如何从原始视频中提取视频片段,同时保留比较完整的视频内容以及如何实现对视频的快速浏览和检索,这就是目前数字视频技术的一个研究热点和难点——视频摘要(Video abstraction).如今国外,如德国曼海姆(Mannhein)大学,美国卡内基梅隆大学(CMU)、明尼苏达州大学,MIT 实验室等,国内,如微软研究院、清华大学、国防科技大学等研究机构都在进行视频摘要的研究[3~8].

1 视频摘要的基本概念

大家知道,一篇文章的摘要,就是对文章的简要

基金项目:十五武器装备预研项目
收稿日期:2001-11-19;**改回日期**:2003-02-21

总结,而视频摘要的概念则是从文本摘要延续而来的,顾名思义,视频摘要就是对一个较长的视频文件的内容所进行的一个简短的小结.文献[3]把视频摘要定义为运动图象的序列,这种定义显然是不全面的,事实上可以根据需要来生成不同抽象层次和形式的视频摘要.应该说,视频摘要是静止图象或者是运动图象的序列(这些图象序列可以附带音频也可以不带),这个序列比原始视频要短很多,但是这个序列应保留原始视频的基本内容,以便能够实现对原始视频进行快速浏览和检索.

视频摘要主要应用在以下领域:

(1) 视频数据的存档和检索　随着多媒体个人电脑和工作站的普及,以及因特网和多媒体数据压缩技术的发展,越来越多的视频信息被数字化存档,由于数据量庞大,检索十分不便,因此需要利用视频摘要技术来改进视频数据的存档.视频摘要是视频数据库的重要索引,因为依靠视频摘要,用户可以快速找到自己感兴趣的视频内容.目前因特网上的视频数据库,在不断完善文本信息的索引的同时,正在积极构建视频摘要的索引.由此可见,视频摘要对于视频的快速浏览和检索是极有意义的.

(2) 影视广告行业的应用　相信很多人有过这样的经历,在电影院里正片即将开播之前,总要播放另一部电影的精彩剪辑(也称为片花),这样的剪辑一般由原始视频中的精彩画面组成,并且不包含故事的结局,这样做是为了吸引观众,为另一部电影作广告宣传.事实上,这是一类比较特殊的视频摘要,它在电影、电视和广告等传媒行业应用广泛.目前,这种视频摘要的制作不仅昂贵,而且耗时费力,但是如果采用较好的自动视频摘要生成系统,那么就可以根据观众的喜好,快速便捷地制作这种电影剪辑.

(3) 家庭娱乐业　一个重要的应用就是视频点播业务,用户可以快速浏览视频摘要,并通过视频摘要来轻松选取自己中意的电影.如果你只记得电影中的一段感人的情节,而不知道片名,视频摘要就会告诉你这部电影的基本内容,并且满足你观看的欲望.

随着视频数据爆炸式地增长以及对视频资源需求的不断提高,视频摘要的应用也将越来越广泛,视频摘要技术也将越来越重要.

2　视频摘要的分类

视频摘要就是通过对视频进行分析处理来自动生成紧凑的能够充分表现视频语义内容的静止或者运动的图象序列.视频摘要还可根据是静止图象,还是运动图象序列分为视频概要(Video summary)和缩略视频(Video skimming)两大类[9],其进一步细分如图 1 所示.

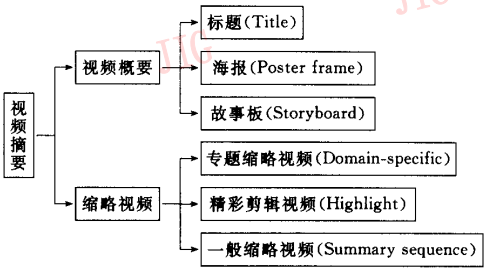

图 1　视频摘要的分类

对于视频概要,可以分为标题、海报和故事板 3 类,其中标题是对视频内容的一段简短的文字描述,这种方式虽是最紧凑最简单的视频摘要形式,但是却很难由计算机来自动生成能准确概括视频内容的文字描述;海报又称为视频代表帧,它是一幅对原始视频有代表意义的图象,它可以提供直观的可视信息,但是无法表现视频的动态特性;故事板是从原始视频中提取的,按照一定顺序和一定形式排列的多帧代表帧图象序列,它可以给用户提供视频的总体描述,在浏览中也可以方便地定位到用户感兴趣的部分.在视频概要的生成过程中,一般不需要伴音和文本的辅助,由于不存在时间同步的问题,因此不仅实现速度快,显示速度也很快.视频概要还可以用全景图拼接法(Mosaic)来表现更加全面和准确的信息,也可以通过一定的空间关系来显示时序图象.

对于缩略视频,可以分为精彩剪辑视频、专题缩略视频和一般缩略视频 3 类,其中精彩剪辑视频就是前面提到的在电影电视中应用广泛的视频摘要,为了吸引观众,剪辑视频一般由原始视频中的精彩画面组成,并且不包含故事的结局;专题缩略视频是特定领域视频的摘要,专题缩略视频的实现一般都要结合该领域的知识来采用比较特殊的方法;一般缩略视频是相对于专题缩略视频而言的,它是一些视频片段的序列,用户可以通过播放这些相对短小的视频片段来了解整个视频的内容.与视频概要相比,缩略视频有其自身的优势,即缩略视频可能比视频概要中单纯的静止图象更加有意义,对用户而言,理解起来更加自然有趣,例如在记录片中,视频的伴

音就包含有重要的信息,因此,在很多情况下,以缩略视频作为摘要更加合适.

3　视频概要的实现技术

视频概要是最能代表视频内容的静止图象集合,因此,关键帧(Key frame)的提取是视频概要实现的主要技术.目前关键帧提取的方法按帧、镜头、场景的视频层次结构划分,主要有基于镜头的和基于场景的关键帧提取方法两类.

3.1　基于镜头的关键帧提取方法

既然镜头被定义为一个连续的视频帧序列,那么在这个序列中就不存在场景或者摄像机运动的突变,因此一个很简单自然的方法就是把每个镜头的第一帧作为关键帧.如果镜头内的内容变化不大,则一帧关键帧就足够了;否则就应该提取多帧关键帧.但是,提取镜头中的哪些帧作为关键帧呢?在目前计算机语义理解还很困难的情况下,大多以低层视觉特性(例如颜色、运动等)为衡量标准来抽取多帧关键帧:

(1) 基于颜色的方法　由于颜色信息不受图象转动的影响,因此基于颜色的关键帧提取方法就被广泛应用,其一种实现方法是在提取镜头的第 1 帧作为关键帧的基础上,再求出接下来的帧与最后一帧之间的颜色信息的差值,如果这个差值超过了一定的阈值,就提取这一帧作为关键帧;另一种是采用聚类的方法,即首先根据颜色直方图的特征,将一个镜头中的所有帧聚类成一定数目的类,然后将提取到的每一类质心位置的帧作为关键帧,但是,基于颜色的关键帧提取方法也存在如下不足:一是帧的选取需依赖于阈值的选择,二是颜色特征还是不能很好地表达视频的语义信息.

(2) 基于运动的方法　这种方法比较适合于时序上有动态变化的帧,其中常用的有光流法.光流法的基本思路是:首先计算镜头中每一帧的光流,然后根据得出的光流场来衡量运动强度,同时分析光流场运动强度函数,最后把具有极小运动强度的帧作为关键帧.

(3) 全景图拼接法　基于颜色和运动的方法都不可能通过选择某一帧关键帧来代表整个视频段的内容,而全景图拼接技术则不仅能将一个镜头中具有相同或者部分相同背景的图象帧连成一幅图象,而且能用一幅图象表示整个镜头的内容,并且从时间和空间上压缩了数据,即排除了连续视频帧时间和空间上的冗余.全景图生成一般分为如下两步:第 1 步对连续帧的运动,套用一定的全局运动模型,例如平移模型、旋转模型、仿射模型、平面透视模型等来进行变换;第 2 步根据估算的摄像机的运动参数来对图象进行变形处理,之后拼接成一幅全景图象.这种全景图拼接法对于背景信息虽然保存得很完整,但是它不能保留前景对象的运动信息,例如,一个镜头表示一个男孩从院子的东端走到院子的西端,全景图就不能包含走的信息.为了解决这个问题,人们把全景图分为静止的背景全景图和前景动态对象的全景图两类,最后把两者结合起来恢复全景图.全景图拼接法的不足在于:它比较适合于包含有确定的摄像机运动的视频片段,而并不适合于真实世界中包含有复杂摄像机运动和频繁背景前景交替的视频.

3.2　基于场景的关键帧提取方法

对于基于镜头的关键帧提取方法,如果是长视频,那么将提取数以百计的关键帧,这样浏览起来不仅费时,而且低效.基于此原因,人们开始考虑基于更高一层的视频单元的关键帧提取法,称为基于场景的关键帧提取法.这里的场景比视频层次结构中的场景更广泛、更丰富,它可以是一幕情景、一个事件,甚至是整个视频序列.

基于场景的关键帧提取方法中比较有名的是 FX Palo Alto 实验室的漫画书(Comic book presentation)[10,11].这种漫画书就是一种特殊的故事板.在该研究项目中,研究人员首先把所有视频帧聚类成预定数目的类;然后根据一段连续帧属于哪一类来对视频进行分割.对于每一分割段,再根据它的长度和出现频率计算一个衡量值,如果这个值小于某一阈值,这个视频分割段就会被忽略;接着提取剩余分割段的关键帧,并通过关键帧的链接可以回放原始视频段.这里,最特别的是关键帧的显示,即比较重要的也就是衡量值较大的关键帧显示较大的图象,而不是很重要也就是衡量值较小的关键帧显示较小的图象,这样即得到一种类似漫画书形式的视频摘要.漫画书中图象帧是从分割的场景中提取的,且在关键帧的显示上也别具特色,即它能从空间顺序上来表示关键帧的重要程度,但是,聚类的数目如何定义,场景的重要程度如何衡量,阈值如何选取,这都是值得进一步深入研究的问题.

除了以上谈到的用关键帧来构造视频概要的方

法,还有很多结合其他技术的视频摘要生成法,如马里兰大学把视频序列表示成高维特征空间的曲面来生成视频摘要.雅典大学把模糊算法和遗传算法(GA)运用到视频摘要中.此外还有结合小波变换、人脸探测等技术来提取关键帧的方法.从目前的发展来看,上面谈到的所有方法都是有利有弊,还没有一种通用的非常有效的方法.

4 缩略视频的实现技术

缩略视频分为视频剪辑和分段序列两类.

4.1 视频剪辑的实现技术

视频剪辑是一类比较特殊的视频摘要,它是原始视频中精彩场景的集合,但是并不包含故事的结局,通俗的称呼是片花.德国的曼海姆大学对剪辑视频曾作过研究[3,12],其研究焦点就是精彩场景的探测和选取.研究人员首先认为包含有强烈对比的前后帧可能包含有重要对象的重要事件;然后他们把表示整个视频段的基本颜色基调的场景也包括在视频摘要中;最后,把所有选取的场景按照时序组织起来,但是,在他们的研究项目中,由于研究人员对问题的复杂性尚考虑不够,所采用的算法还比较简单,因此效果有时候不是很好,还有待进一步提高.

4.2 专题缩略视频的实现技术

专题缩略视频是一种针对某一特定领域视频数据的缩略视频.对于专题缩略视频,一般可结合该领域的专题知识,采用特殊的方法来生成视频摘要.文献[13]设计了一种专门针对该研究机构每周例会的视频摘要系统,即利用例会比较统一的履行程序,把低层的信号事件和高层的语义事件关联起来生成缩略视频.可见,专题缩略视频是从专题知识出发,更多的是采用基于模型,而不是基于内容的方法来生成摘要.

4.3 一般缩略视频的实现技术

事实上,选取整个视频中最精彩的图象帧往往是由人主观确定的,而且如何把人的认识与计算机匹配起来是一件非常困难的事情,基于以上原因,目前缩略视频的重点集中在一般缩略视频的研究.一般缩略视频实现的一个最直观的方法就是通过压缩原始视频来加速视频回放的速度.微软研究院采用时域压缩技术(time compression technology)来获取视频片段,这种方法虽然有一定的效果,但是它存在压缩比的限制,因为这些压缩算法是依赖于语音速度的,如果压缩比过高的话,那么语音将无法理解.

从目前视频摘要技术的发展来看,一般缩略视频的实现主要采用多特征融合[6,14]的方法,也就是结合文本、音频和视频等媒体的特征来生成视频摘要.其中比较有名的是卡内基梅隆大学的研究[15,16],文献[15]、[16]中,研究人员致力于从原始视频中提取最有代表意义的音频和视频信息,以创建一段简单的缩略视频,即首先从一些文字说明中提取关键字,同时从视频中探测字幕;然后根据这些关键字创建语音摘要;接下来就是抽取代表帧,主要包括以下几类:包含有关键人脸或者文本的帧、表明摄像机运动的帧、视频场景中的开头帧,由于以上这几类图象帧提取的优先级是依次降低的,因此这些提取出来的帧不一定按照时间顺序排列,但是从视觉效果上讲,这样的缩略视频更加合适;最后,按照文本、音频和帧的匹配关系来生成缩略视频.这种方法对于那种有附加语音或者文本信息的视频非常有效,例如记录片等,而对其他的视频效果则不是很好.

5 总结及展望

综上所述,目前视频摘要技术具有如下一些特点:

(1) 目前的视频摘要技术的研究重点主要集中在低层次上,尽管人们企望能从更高一层的视频结构(例如场景)来对视频进行分析,但是目前任何层次上的研究都是不完善的.尽管物理特征是分析识别图象的基础,但是如何把高层概念和低层特征关联起来呢?如在日常生活中的“花、房屋、海滩、日出”等概念都属于多媒体数据的高层语义,过去对基于低层特征的检索已作了大量的工作,如果能够建立起这些低层特征与高层语义概念的关联,就能够用计算机来自动提取媒体数据的语义.在特定应用领域中,低层特征与高层语义概念的关联,虽可能容易一些,但是对于一般性的特征,企望通过突破低层特征的壁垒来获取高层语义是非常困难的.

(2) 视频分析的最终目标就是让计算机视觉达到或接近人的视觉水平,但目前计算机的视觉水平与人的视觉能力相差非常远,仍存在着质的区别,因为计算机视觉研究的进展是逐步的,需依赖于各个方面理论的发展,不能够期望短时间就可以突破.据分析,从人的视觉原理出发去分析计算机视觉可能是一个有效的途径.

(3) 视频摘要的有效实现不仅需依赖于计算机

视觉的研究，而且需依赖于知识库及基于知识的联想与判断等人工智能技术的发展.

（4）由于视频内容的复杂性和人类理解的多样性，以及语义理解的屏障，因此生成完整、准确、令人满意的视频摘要是比较困难的. 在目前的研究基础下，如果不能从语义的角度去分析理解视频，就不能得到很好效果的视频摘要.

参 考 文 献

1 Ngo Chong-Wah, Zhng Hong-Jiang, Pong Ting-Chuen. Recent advances in content based video analysis [J]. International Journal of Image and Graphics, 2001,1(3):445～468.

2 Atsuo Yoshitaka, Tadao Ichikawa. A survey on content-based retrieval for multimedia databases [J]. IEEE Transactions on Knowledge and Data Engineering, 1999,11(1):81～93.

3 Lienhart R, Pfeiffer S, Effelsberg W. Video abstracting [J]. Communications of ACM[J], 1997,40(12):55～62.

4 Wachtlar H, Christel M, Gong Y *et al*. Lessons leaned from the creation and deployment of a terabyte digital video library [J]. IEEE Transactions on Computer,1999,32(2):66～73.

5 Hari Sundaram. Segmentation, structure detection and summarization of multimedia sequences [D]. New York,USA: Columbia University, 2002.

6 王辰. 多媒体融合分析技术的研究与实现[D]. 长沙：国防科学技术大学，2002.

7 熊华. 视频内容结构化技术的研究与实现[D]. 长沙：国防科学技术大学，2001.

8 曹莉华. 视频媒体的基于内容处理和检索的研究与实现[D]. 国防科学技术大学，1998.

9 Li Ying, Zhang Tong, Daniel Tretter. An overview of video abstraction techniques [R/OL]. In: HP Corp Technology Report, HPL-2001-191, 20010809, External. http://www.hpl.hp.com/techreports/2001/HPL-2001-191.html, 2001.

10 Boreczky J, Girgensohn A, Golovchinsky G *et al*. An interactive comic book presentation for exploring video[A]. In:Proceedings of Computer Human Interaction 2000[C], New York, USA: Association for Computing Machinery Press,2000:185～192.

11 Uchihashi S, Foote J, Girgensohn A *et al*. Video mange: Generating semantically meaningly video summaries [A]. In: Proceedings of Association for Computing Machinery, Multimedia'99[C], Orlando, Florida, USA, 1999:383～392.

12 Lienhart R. Dynamic video summarization of home video[A]. In: Proceedings of SPIE Storage and Retrieval for Media Database 2000[C], San Jose,CA,USA,2000,3972:378～389.

13 Russell D M. A design pattern-based video summarization technique: moving from low-level signals to high-level structure [A]. In: Proceedings of the 33rd Hawaii International Conference on System Sciences[C], Maui, Hawaii, USA, 2000, 3:3048.

14 Hu Jian-ying, Zhong Jia-lin, Amit Bagga. Combined-media video tracking for summarization [A]. In: Proceedings of Association for Computing Machinery, Multimedia'01[C], New York,USA:ACM Press, 2001:502～505.

15 Christel M G, Winkler D B, Taylor C R. Multimedia abstraction for a digital video library [C]. In: Proceedings of the 2nd Association for Computing Machinery, International Conference on Digital Libraries[C], Philadelphia, Penn. USA, 1997: 21～29.

16 Micharl G C, Alexander G H, Adriennce S W *et al*. Adjustable filmstrips and skims as abstractions for a digital video library [A]. In: Proceedings of IEEE Forum on Research and Technology Advances in Digital Libraries [C], Baltimore, Margland,USA:IEEE Press, 1999:98～104.

陈剑赟　1977 年生，1999 年获国防科技大学系统工程与数学系信息系统工程专业学士学位，现为国防科技大学多媒体研究与开发中心硕博连读生. 主要研究方向为视频内容分析、视频摘要.

老松扬　1968 年生，1996 年获国防科技大学系统工程专业博士学位，副教授，中国计算机学会 CSCW 专委会委员. 主要从事基于内容检索、多媒体数据库的研究.

吴玲达　1962 年生，教授，博士生导师，中国计算机学会多媒体专委会常务委员、中国计算机学会虚拟现实与可视化专委会常务委员. 主要从事多媒体信息系统和虚拟现实技术的研究.