●何立民 万跃华

数字图书馆中基于内容的视频检索关键技术

摘 要 实现数字图书馆中基于内容的视频检索的关键技术包括:视频数据建模、视频分割、视频索引、视频查询与浏览。有待深入研究的问题有:算法的引进、视频的检索反馈等。图 1。参考文献 21。

关键词 数字图书馆 基于内容的视频检索 多媒体信息检索
分类号 G250.76

ABSTRACT Key techniques for the realization of content-based video retrieval in digital library include video data modeling, video segmentation, video indexing, and video searching and browsing. Problems to be solved include the improvement of algorithms, the feedback of video retrieval, etc. 1 fig. 21 refs.

KEY WORDS Digital library. Content-based video retrieval. Multimedia information retrieval.
CLASS NUMBER G250.76

如何有效地组织、管理和检索大规模的视频数字图书馆,是海量信息处理的瓶颈,已成为国内外研究热点。而基于内容的检索技术便是解决这一问题的关键技术之一[1]。

1 数字图书馆中基于内容的视频检索关键技术

基于内容的视频检索(Content-Based Video Retrieval),是一种新的检索技术。它能从数据库中查找到具有指定特征或含有特定内容的图像(包括视频片段)。它区别于传统的基于关键字的检索手段,融合了图像处理、模式识别、计算机视觉、图像理解等技术,具有如下特点:(1)直接从视频数据中提取信息线索。(2)它是一种近似匹配,与常规数据库检索的精确匹配方法明显不同。(3)自动提取并描述视频的特征和内容。如图 1 所示,首先要构造视频结构,将视频序列分割为镜头,并在镜头内选择关键帧,这是实现一个高效的基于内容的视频检索系统的基础和关键。然后提取镜头的特征以及关键帧的视觉特征,作为一种检索机制存入视频数据库。最后根据用户提交的查询按照一定的特征进行视频检索,将检索结果按相似程度提交给用户。特征的提取和检索算法的优劣决定了整个检索系统的效率和性能。基于内容视频检索技术已经在数字图书馆中得到广泛应用。实现数字图书馆基于内容的视频检索技术的主要关键技术包括以下几个方面。

1.1 视频建模

要进行基于内容的视频检索,首先要建立一个合理的视频数据模型。

视频数据由于其本身的综合性(包含声、视内容及高级语义内容)、结构的复杂性(非格式化)及具有时空多维结构,要用一个恰当的数据模型把现实世界的视频反映到信息世界及机器世界,问题十分复杂。传统的数据库模型(如关系模型)无法直接应用于视频数据库,即使作某些改进补充,也难以适应视频数据这类复杂的数据类型。

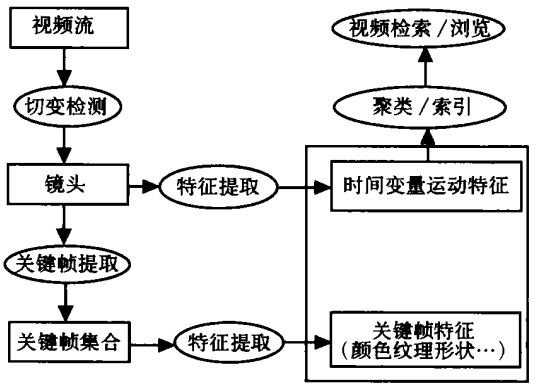

图 1 基于内容的视频检索系统框架

传统的视频数据模型只反映空间信息的表达式上增加时间因素[2~4]。这类模型并没有反映视频数据的一些主要特征如空间特征及视频语义内容,也没有反映视频原始素材的共享及视频数据的独立性。但它们的一些重要概念(如时态区间、区间关系

等)成为构成各类视频模型的重要内容。

在视频模型建立中,Ron 等人引入了视频单元的合成及运算体系,采用多种代数方法,形成基于代数的视频数据模型[5]。在视频单元合成成为新的视频流时,这些视频单元的排列会形成新的上下文,产生新的语义,在视频数据模型中也应反映这种上下文关系。另外在视频数据建模中还引入了面向对象思想[6]。

由于视频数据本身的复杂性及应用的广泛,要建立统一的能广泛应于多个领域的普遍性模型还有困难。受目前图像理解、计算机视觉、人工智能等学科发展水平的限制,视频数据自动地分段及抽取视频的高级语义特征也还存在不少困难。目前还不应对视频数据模型提出过高的不切实际的要求,而应以建立有限自动化且应用于某些特定领域的模型为目标。

1.2　视频分割

对视频的处理主要包括视频分割、代表帧的抽取及视频特征的提取等。视频分割是最主要的一步。视频分割是指对图像或视频序列按一定的标准分割成区域,目的是为了从视频序列中分离出有一定意义的实体,这种有意义的实体在数字视频中称为视频对象。视频分割主要有两种方法:数据驱动方法和模型驱动方法。

视频分割按用途大致可分为用于编码目的和基于内容可操作两大类。前者一般基于视频图像低层次级(像素级)的特征,后者要依靠视频图像高层次级(对象级)的特征。按照人工参与的程度,通常分为自动分割和半自动分割技术。

1.2.1　自动分割方案

视频分割技术是在静态图像分割的技术基础上发展起来的。静态图像的分割算法一般是利用图像上颜色、灰度、边缘、纹理等空间聚类信息进行基于区域的分割,典型的可分为单层次方法和多层次方法。在单层次方法中,传统上有基于边缘图的方法、K- 最近邻域法等。多层次方法在研究领域越来越得到重视,如分裂合并、金字塔链接和形态学方法等[7~9]。使用形态学滤波器和分水岭算法的形态学分割,由于算法效率高,越来越多地得到应用。然而进行视频分割时,这些空间分割方法没有利用视频序列在时间轴上的信息,计算量大且不能得到令人满意的结果。通常同时利用视频图像对在空间和时间轴上的信息进行分割。自动分割算法大致可分为基于光流法的分割、运动跟踪法和基于变化区域检测的时空法 3 种。

1.2.2　半自动分割方案

在面向对象可操纵的应用中,需要从视频序列中提取出人感兴趣的视频对象,即语义视频对象。当前的自动分割技术通常是根据对象在空间或运动方面的聚类信息进行分割,但是语义视频对象往往具有几种不同的颜色、灰度和运动,因此采用自动分割方法很难实现。既然语义视频对象是人主观上定义的,那么在现阶段,人凭自己的认知参与视频分割过程从某种意义上来说是不可避免的。在不要求实时性但对视频对象边界精度要求较高的应用场合,可以采用人工交互的方式确定分割对象,提高视频分割的精度。半自动分割的一般做法是人通过图形用户界面(GUI)对视频图像进行初始分割,对后继帧的分割则采取自动分割算法。目前较为成功的半自动分割算法主要有:按被分割对象的性质进行跟踪,基于变化检测的方案,基于形态学算法的方案。这些半自动分割算法尽管取得了良好的分割效果,最大的缺点是需要人工参与,不能用在实时性的应用中。

虽然人的眼睛可以轻易地分辨出视频对象,但依靠计算机提取视频对象的技术还不成熟。衡量一个分割算法优劣主要考虑:分割质量 ——在尽可能减少人工参与的情况下,能够得到视频对象的精确边界;计算机复杂度 ——计算量小,无须依赖于高档的计算机;算法通用性 ——能在对视频对象的颜色、形状、运动和类型没有先验知识的情况下分割视频对象。

半自动分割虽然分割质量较好,但依赖于人工的交互,不具有实时性。而当前的自动分割算法都是在一些图像的低层次的特征上进行的,大多面向特定的应用,通用性差,分割质量也难以让人满意。另外,绝大多数的自动分割方法都需要用户调节某些参数,从某种意义上来说,这也是一种人工交互。对于特定应用和场景不变的系统,用户可以根据先验知识指定这些参数,但不具有通用性。自动判断这些参数的算法正处于研究阶段,要成功地对实时视频序列进行自动分割需要在人工智能、图像理解、运动分析和人眼生理特性等方面做进一步研究。

1.3　视频索引

视频数据包含极其丰富的语义内容,结构复杂多样,在物理层次上,视频是二维像素阵列的时间序列,与语义内容并不直接相关。要实现基于内容的视频检索,必须突破传统的基于一个或多个关键域(或属性)建立索引和表达式检索的局限,直接对视频内容进行分析,抽取特征和语义,并利用这些内容

特征建立索引。

视频索引从不同的角度出发有不同的分类方式,从选取的索引内容出发,可以分成 3 类:基于注释的索引(Annotation-based Indexing),基于特征的索引(Feature-based Indexing)和基于特定领域的索引(Domain-specific Indexing)。

1.3.1　基于注释的索引

所谓视频注释,是指用一些描述性的信息(如文字、声音或图形)来表述所指向的视频段。目前基于注释的索引技术的研究主要集中在注释语言的选择、注释结构的设计,以及方便的人机交互式注释界面的设计 3 个方面[10]。

注释语言可以分为自然语句、关键词、源注释等。自然语句注释可以充分表达视频数据的语义内容,但随意性相当大。关键词注释可以预先给出供选择的关键词以减少随意性,但也存在信息表述不完全等不足,它的缺点在于忽视了多媒体与文本在数据组织方式、数据量等方面都有很大差别。而且关键词查询的方式在视频应用领域有两个致命的缺陷:首先,人工注解需要大量劳动力;其次,不同的人对同一视频内容有不同的理解,这种理解上的主观性和注释的不精确性会引起检索过程中匹配误差。源注释的基本思想是利用数字摄影机在视频数据流中加入相关信息作为视频注释的依据。

1.3.2　基于特征的索引

视频数据包含语法内容和语义内容,基于特征的索引模式正是利用可自动识别的语法内容建立部分视频索引,基于特征索引技术的目标是全自动索引,这些技术主要依赖于图像处理算法分段视频,识别关键帧,并从视频数据中抽取出关键特征。根据这些关键特征,建立索引。关键特征可以是颜色、纹理、运动对象等。

基于特征的索引技术的研究目前主要集中在图像特征索引和视频特征索引。不带时间延展性的特征称为图像特征如颜色、纹理、轮廓、形状等。视频特征表示将在镜头层次表示视频数据的时间特征。视频除了具有一般静态图像的特征,还具有动态特征。目前视频特征提取的主要任务是:从图像序列中检测出运动信息、识别与跟踪运动目标和估计三维运动和结构参数。由于一旦检测出运动信息或估计出三维运动和结构参数,可以通过目标的形状、运动等属性,根据目标模型的先验知识,识别出目标,就能知道并预报目标的当前与未来位置,也就解决了目标的识别和跟踪问题[11,12]。

1.3.3　元特征索引

元特征是指有关视频数据的一些基本特征,如视频的出品人或公司信息、导演、出品日期、视频文件长度、原始载体、版权认证号、类别(剧情类、非剧情类等)、压缩方式(如 MPEG □等)、制式(PAL NTSC 等),甚至演员等。元特征是与视频内容无关的特征,它一般应用于视频数据整体而较少用于视频分段,故也称为基本特征。元特征的输入通常是在一段新的视频数据插入视频数据库时由人工或半人工输入完成。某些嵌入视频中的文本内容如字幕、演员表等提供了有关视频的重要的元(meta)特征可用光学特征识别技术(OCR)来提取[13]。

1.3.4　基于领域知识的索引

这是指专门针对某个应用领域建立的索引,它们一般有固有的模式,例如新闻视频分主持人段和新闻内容段,篮球比赛分节等。对特定领域的视频索引,可以首先根据它们的固有模式建立逻辑(高级)视频结构模型,在索引建立过程中,在对视频数据进行特征提取和分析的基础上将结果与模型匹配。一旦确定了匹配关系,就可以将模型对应的语义赋给相应的视频数据单元。

目前,在视频数据库的应用中都是选择具有良好逻辑结构的视频单元来进行索引。由于对视频数据库的研究远没有文本数据库那样成熟,上述 3 种索引方法都各有利弊。如何建立高效、实用、方便的索引仍值得研究。

1.4　视频数据查询与浏览

常用的视频查询方法是通过特定的查询语言或通过可视事例方式来完成。用户要查找一个对象时,可以用查询语言或事例形成一个查询条件。系统把查询条件中描述的特征转化为具体的特征矢量,或对事例进行特征提取。将查询描述的特征与特征库中的特征按照一定的匹配算法进行相似度计算,并返回一组满足一定相似度要求的候选结果。对系统返回的查询结果,用户可以通过浏览来挑选,直至得到满意结果。或者从候选结果中选择一个示例,经过特征调整后,又形成一个新的查询条件。这样不断重复操作,直到用户对查询的结果满意。

目前视频查询方式可分为按查询内容、匹配精度要求、数据单元大小等进行分类。按查询层次分类有语义信息查询、元信息查询、视听信息查询。按提交查询形式分类有图像查询、运动特征查询、声音

查询、文本查询、几何查询。按查询功能分类有定位查询、浏览查询、跟踪查询。按查询匹配精度分类有精确匹配查询、相似匹配查询。按查询的数据单元大小分类有基于帧的查询、基于视频片段的查询、基于视频流整体的查询。按查询的表达方式分类有直接查询、示例查询、渐进查询。

由于视频数据查询的多样性和复杂性,传统的查询语言如 SQL 已无法满足要求,新的视频语言有 TSQL、STL、VSQL 等。

对于视频来讲,浏览与有明确目的的检索是同样重要的,当用户对所要检索的目标并不十分明确时,往往需要对视频数据进行快速浏览以便寻找感兴趣的内容。目前视频浏览采用分层结构和集束分类等方法,已成为基于内容视频检索的新的重要研究方向。

2　基于内容的视频检索原型系统与视频数字图书馆系统

QBIC(Query By Image Content)是 IBM Almaden 研究中心开发的第一个商用基于内容的图像及视频检索系统。系统中的基本元素是由图像子集构成的场景、一系列连续帧分割而成的镜头及运动对象。它提供了对静止图像及视频信息基于内容的检索手段。它的系统结构及所用技术对后来的视频检索有深远的影响。QBIC 还是少数几个考虑了高维特征索引的系统。查询结果可以按照相关的序列指导子序列继续查询。这种方法能够使用户更加快速和简捷地对可视化信息进行筛选与确定[14]。

JACOB 基于内容的视频检索系统,可进行视频自动发段并从中抽取代表帧,并可按彩色及纹理特征以代表帧描述基于内容的检索[15]。

VisualSEEK 图像查询系统和 WebSEEK 图像及视频搜索引擎是美国哥伦比亚大学开发的基于内容检索原型系统[16]。该系统的主要特点是用到了图像区域的空间关系查询和直接从压缩数据中提取视觉特征。所用到的视觉特征有颜色集、纹理特征的小波变换。为了加快检索过程,还开发了基于二叉数的索引算法。

卡内基-梅隆大学的 Informedia 数字视频图书馆系统,结合语音识别、视频分析和文本检索技术,支持 2000 小时的视频广播的检索;实现全内容的、基于知识的查询和检索。它综合应用了图像处理、语音识别、自然语言理解、视频分析的最新技术,展示了计算机多媒体信息处理的无限空间[17]。它已经在该大学及当地学校和英国开放大学使用。

美国堪萨斯大学的数字视频图书馆系统(DVLS)的目标是存储、索引及检索声视频信息并通过因特网及国家信息基础设施实现视频共享的技术,已建立了一个称为 VISION 的原型系统及一个视频数据库[18],数据库中包含了有 1000 多小时的由多个广播通讯公司提供的视频信息。

此外还有许多优秀的原型系统,并在一些领域得到应用。如 OVID 是一个采用面向对象技术的视频对象数据库系统[19]。系统中建立了基于面向对象技术的浏览及查询机制并设计了视频查询语言 VideoSQL。VideoSTAR 是一个利用通用视频数据库框架模型建立起来的一个视频存储及检索实验系统,系统支持视频数据的共享及各种商用数据库的连接使用[20]。

3　有待深入研究的问题及进一步研究方向

3.1　算法有待改进

由于视频的数据量大,处理时间长,算法处理的速度很重要(由于大量的视频数据是以压缩形式存放的,直接对压缩数据进行处理可以节约时间)。

镜头的检测直接关系到漏检和误检率,而且镜头的检测所用的颜色、纹理、运动等特征可以用于最后的检索处理,所以镜头的检测算法应进一步作重点研究。

目前在切变检测和检索算法中底层语义特征比较成熟,检测结果比较理想,故用得最多。但高层语义,如用户感兴趣的事件、运动、物体特征的变化等,也是用户的检索需求,这也是视频检索算法方向之一。

3.2　阈值的选取

阈值选取不当会造成误检和漏检。有的视频变化缓慢,应选取较小的阈值;反之则应选取较大的阈值。应不断试验,尽量达到均衡,并综合利用人的知识进行人机交互式学习选取合适的阈值。以上所介绍的各种方法多受阈值选取优劣的限制,近年来趋向于更加鲁棒性的研究,如利用 K 均值法的检测,可以减少阈值选取的限制。

3.3　检索效果的评价尚没有标准

视频检索效果评价主要使用的是查全率和查准率两个指标。用户在评价算法的时候,可以预先选定含有特定目标的视频作为一组相关的视频,然后根据返回的结果计算查全率和查准率。查全率和查准率越高,说明该检索算法的效果越好。

3.4　视频的检索反馈

基于内容的视频检索系统中,最常用的检索方式是例子视频查询,即用户提交一部视频,系统返回相似的一系列视频,但怎样定义的两部视频是相似的,仍然是困难的问题,限制了检索系统的应用范围[21]。而且由于视频内容的复杂性,不同用户在检索过程中,即使对同一部视频,其注重的角度也有可能不同,因此接受用户的反馈意见,当用户对查询结果不满意时可以优化查询结果,突出用户的需要,仍需要进一步深入研究。

3.5　视频多特征的综合检索方法

基于内容视频检索还要解决多种检索手段相结合的问题,以提高检索的效率。对于单一特征检索手段,由于其约束信息不足,在返回目标视频的同时往往会返回大量其他也满足此检索要求的视频。采用多个检索手段相结合的方法无疑可提供更多的约束而使得返回视频中目标视频的比率得到提高,但检索手段间的融合是所要解决的问题。MPEG7 标准,其目标就是实现集高层语义特征和低层视觉特征的基于内容的多特征综合检索,今后研究的热点之一将是高层的基于语义内容的视频检索。

此外,应以认知科学的研究成果分析视频内容的特征和人对视频的认知。视频信息在人脑中的长期记忆为心像,人对心像的记忆、检索等操作过程实际上是形象思维过程,因此形象思维科学中关于心像的表征和计算模型将对基于内容的视频检索提供一定的指导。今后研究的热点之一将是视频序列图像中人的行为识别和分析。

综上所述,可以看出基于内容的视频检索仍然是一个开放性的研究课题,其研究将涉及认知科学、人工智能、模式识别、图像处理、知识库系统、计算机图形学、数据库管理系统、用户模型、信息检索等多个领域。许多技术还处在实验阶段,许多问题有待解决。

参考文献

1　卢汉清,孔维新,廖明等. 基于内容的视频信号与图像库检索中的图像技术. 自动化学报,2001,27(1)
2　T. D. C. Little , Time-Based Media Representation and Deliver. In ACM Press and Addison-Wesley Publishing Company , 1994 , Multimedia System , chapter 7
3　T. D. C. Little , et al. A Digital On-Demand Video Service Supporting Content-Based Queries , In Proceedings of ACM Multimedia 1993
4　T. D. C. Little , et al. Synchronization and storage model for Multimedia Objects , IEEE Journal on Selected Areas in Communications. April 1990
5　Ron Weiss et al. Content-based access to algebraic video. In IEEE International Conference Multimedia Communications , 1990
6,19　Eitetsu O , Katsumi T. OVID: Design and implementation of a video-object database system. IEEE Transactions on Knowledge and Data Engineering , 1993 , 5(4)
7　Horowitz S L , Pavlidis T. Picture segmentation by a tree traversal algorithm. Journal of the ACM , 1976 , 23(2)
8　Burt P J , Hong T H , Rosenfeld A. Segmentation and estimation of image region properties through cooperative hierarchical computation. IEEE Trans on Systems , Man and Cybernetics , 1981 , 11(12)
9　Salembia P , Torres L , Meyer F et al. Region-based video coding using mathematical morphology. Proc of the IEEE , 1995 , 83(6)
10　郑鹏,李订方,刘清青. 基于注释的视频索引. 计算机应用,1999,19(7)
11　Petkovic M. et al. Content-based video indexing for the support of digital library search. Proceedings 18th International Conference on Data Engineering , 2002 ,494 - 5
12　何立民,万跃华. 数字图书馆中基于内容的图像检索技术. 现代图书情报技术,2002 年刊
13　Sato T et al. Video OCR: Indexing digital news libraries by recognition of superimposed captions. Multimedia Systems , 1999 , 7(5)
14　Niblack W , Zhu XM , Hafner JL , Breuel T et al. Updates to the QBIC system. Storage and Retrieval for Image and Video Databases , 1997 , VI
15　Maroo La Cascia , Edoardo Ardizzone. JACOB: Just a Content-Based Query System for Video Database , Proc ICASSP-96 , May 7 - 10 , Atlanta , GA
16　Smith J. R , Chang. S-F. VisualSEEK: a fully automated content-based image query system. In Proc. ACM Intern. Conf. Multimedia , Boston , MA , 1996
17　Christel , Michael G. et al. Interactive maps for a digital video library. IEEE Multimedia , 2000 , 7 (1)
18　Gauch , Susan , Li , Wei , Gauch , John. Vision digital video library. Information Processing & Management , 1997 ,33(4)
20　Rune H , Roger M. Searching and browsing a shared video database. Proceedings of the International Workshop on Multi-Media Database Management Systems , 1995
21　吴翌,庄越挺,潘云鹤. 视频的检索反馈. 计算机研究与发展,2001,38(5)

何立民　浙江工业大学图书馆馆长。通讯地址:杭州市。邮编 310032。
万跃华　浙江工业大学图书馆信息咨询部主任。通讯地址同上。
(来稿时间:2002-05-27)