

基于内容的视频/多媒体数据库检索技术

Content-based Retrieval Technique in Video/Multimedia Database

陈慧琼 陈慧君 刘宗田

(上海大学计算机学院 上海200072)

Abstract While large video/multimedia databases became increasing common, traditional retrieval methods are no longer adequate for video data. Since video information is temporal, spatial, unstructured and massive, the new content-based video retrieval, which can retrieve databases by the video content, will supply user better. We can classify Video CBR into two categories: the improved traditional methods and the visual methods. In this paper we describe the main idea and recent studies on video CBR. After analyzing different development models of some typical video CBR systems, we point out their distinguished implement characters and shortcomings. At last, we suggest some issues that will improve query efficiency based on forecasting the trend of video retrieval technique.

Keywords Video retrieval, Content-based, Multimedia database

1. 引言

在当今的信息社会里,多媒体技术、计算机网络技术的迅速发展使得多媒体技术的应用也随之迅速发展。然而,可视化信息的大量生产和传播也带来一个问题,即面对信息量如此浩大的多媒体数据库,用户应如何找到自己需要的信息。例如:用户可能提出这样的要求,“找出数据库中所有赛车的场面”。我们需要研究的是,信息检索系统应如何适当表达这个要求的内容,并在数据库所有的视频信息中找到符合这个查询要求的信息返回给用户。于是,大规模多媒体数据库的索引和检索具有的巨大实用价值使得它在国外已成为一个重要的科研课题。

多媒体检索一般分基于表示形式和基于内容两种类型:前者与数据类型和数据结构有关,不需要对多媒体内容做任何分析,实际应用中常用于复合媒体的检索,例如“找出有声音注释的图像”就属于这一类;而基于内容的检索则根据媒体的内容及语义来进行查找。基于内容检索中最容易实现的是查找文本媒体内容,并且现在已有了有效的文字检索系统,这些系统将待查询对象与文本内容进行比较和匹配,并允许进行相似检索,如“找出1999年全年在中文期刊上发表的与计算机相关的论文的情况”。这个查询中有些部分是精确的,如“1999年全年”、“发表在中文期刊上”;有些部分却是不精确的,如“与计算机相关”的论文。一篇文章是否与计算机有关,不同的人看法可能不同,这样该文章是否应被检索到就是有争议的。通常检索系统通过表示实体的特性来计算其相似性,然后用一个距离函数评价一个实体与另一个实体的相似程度,最后把检索到的文章(即相似系数高于某一相关系数)都列出来,这种方法可较好地解决文本的相似性检索问题。但对于图像、视频、音频等类型的媒体来说,基于内容的检索就困难多了,这类检索的关键是语义的理解和表达问题,如前面提到的查询“赛车场面”,虽然简单,却很难用现有检索方法有效地检索出来。目前还没有一个令人满意的有效的多媒体信息检索系统。

一直以来,多媒体检索都是并还将继续是以图像和视频

为主导。虽然语音识别已发展得比较成熟,但很少有讨论涉及如何检索单独音频媒体的内容。因此,本文在讨论多媒体数据库基于内容的检索时,主要关心的就是视频数据库的检索技术。

2. 用传统方法进行视频检索的问题

和传统数据相比,视频信息具有以下特点:数据量大;信息非结构化;信息长度不固定;具有时态特性,这些数据特征引发了检索中的以下问题:

(1)信息内容的提取 所谓信息,指的是数据中提取出来的对数据的解释,它需要根据特定的上下文并利用知识才能推导解释出来,因此同样的数据可以提供不同的信息。对于可视化信息来说,它的内容表示依赖于人对该可视信息的感知力和理解力,显然,对于同样的视频,不同的人由于理解不同可能对它有多种不同解释,这种解释多义性使我们很难从视频中自动生成适合每个人需要的内容信息。

(2)查询的表达 在传统数据库的检索中,查询一般由文字和数字通过关系代数描述表达出来,这些描述和数据库中数据的属性权值有简单的对照关系。但多媒体或视频数据库中,可以提取出的信息内容的类型多种多样,如视频的颜色、纹理、运动、语义等,都可以作为被提取的信息内容,当用户需要查询这些内容时,查询条件若用文本或数字来表达就未必适当,而如何表达具有时态特征的查询也是需要解决的问题。

(3)相似性描述和度量 视频信息的非结构化使得我们很难像处理文本相似性一样,用一个简单的度量函数来比较两段视频间的相似性。但若用文字来描述视频信息并作为相似性比较的依据,则需要考虑提取的合适性,如何用适当的方式表达视频内容,如何在数据库中通过相似性比较搜索出相应的视频片段等问题都是难以解决的。

(4)时态问题 时态性是视频数据的一个重要特征,无法用传统数据库中的常规数据结构有效表示。而在基于内容的视频检索中,用户可能提出包含时态关系的查询,如“找出A和B同时出现的片段”。另外,在检索中,系统提供给用户的检索结果不应仅仅是数据库中存储的包含该内容的独立长视

陈慧琼 助教,研究方向:多媒体数据库。陈慧君 研究生。刘宗田 教授,博导,研究方向:软件工程,数据库知识挖掘。

视频段,还应检索并定位到视频段内部包含该内容的较短视频序列上,因此必须扩充数据模型的时态表达能力和对时态的检索能力。

3. 基于内容检索的基本概念

视频数据的上述特性使得无法通过简单的匹配来达到视频检索的目的,传统的查询方法不能有效查询视频/多媒体数据库。为了使数据库系统能对声音、图像等媒体进行内容语义分析,并达到更深的检索层次,人们引入了基于内容的多媒体/视频数据库查询技术。

所谓基于内容的检索,是指根据媒体对象的语义和上下文联系进行检索,它有如下特点^[1]:

(1)从媒体内容中提取信息线索;基于内容的检索突破了传统文件检索的局限,直接对图像、视频、音频进行分析、抽取特征,使得检索工作更加接近媒体对象;

(2)提取特征的方法多种多样;

(3)人机交互;人对于特征比较敏感,能迅速分辨目标的轮廓等特征,但对于大量的对象,一方面人们难以记住这些特征,同时人工从大量数据中查找目标非常费时,又是一种重复性工作,而这些正是计算机的长处。因此人与计算机间应相互合作进行检索。

(4)检索过程是一种近似匹配;这与传统数据库的精确匹配方法有明显不同。

而 S. F. Chang 等在文[2]中认为,一个先进的基于内容的视频检索系统应具有以下的特征:自动视频分割和跟踪;包含颜色、纹理、形状、运动的丰富的可视化特征库;能进行复合对象的查询;查询的空间-时间约束;在 WWW 上的交互查询和浏览。

4. 基于内容的视频检索方法

目前基于内容的检索方法主要分为两种:在传统查询的基础上扩充对视频的表达和检索能力,以及根据视频可视化特征而提出的可视化查询方法。

1. 空间-时态关系描述法

空间、时态关系是视频等多媒体数据固有的重要特性,但传统数据模型本身并不能有效表示和管理它们,因此在视频检索中需要扩展它的时间-空间关系表达能力。

信息构成成分之间的空间关系,一个直接的表达方式是用直角坐标表达。组成对象的空间位置由坐标表达,而各组成对象间的关系可以用数学方法计算出来,这种方式要求空间关系表达精确;另一种表达方式则改善了抽象空间关系的表达和索引,这种索引技术运用了二维字符串来表达图像中的对象空间关系,二维字符串代表了这些对象的抽象位置,包括它们在垂直和水平方向上的顺序。此外,还有一种方法是定义一个二元关系集合作为表达对象空间关系的初始值,包括“左”、“右”、“前”、“后”、“上”、“下”、“内”、“外”和“交迭”^[3]。二元集合法和二维字符串表达一样,适于空间关系的粗略估计。查询时可以用这两种表示方法表达对象间的空间关系。

时态关系的表达方法有两种:基于点的表达,以及基于间隔的表达。基于点的表达用时间线上的点来指示对象位置;基于间隔的表达利用对象发生的时间间隔来说明对象间的时态关系,目前大部分的研究集中在基于间隔的表达方法上。基于组成对象表达的时间 Petri 网考虑了组成对象的时态顺序表达,同时讨论了基于间隔的时态关系:“前”、“完全相等”、“相

接”、“交迭”、“期间”、“开始”及“结束”^[4],这个基本二元集合还可以延伸为 n 元时态关系。基于间隔的时态关系接近视频帧情节的编辑/构建,比基于点的表达更符合人类认知,因此在视频数据库中,基于间隔的表达更适合表达对象间时态关系;在检索时,我们可以通过基于间隔的方法用时态关系集合来表达待查询对象间的时态关系。

2. 可视化查询方法

可视化方法突破了传统方法只能运用文本或数值形式表达和检索的局限,能可视化地给出查询描述,这使得用户能直观地表达查询要求,也令查询要求和待检索数据间更为接近。可视化查询方法主要有两种类型:范例查询法和草图描绘法。

(1)范例查询法 通过浏览等方法,选择库中的某个媒体作为例子。颜色、纹理、形状均可以作为示例进行查询,系统将从用户的示例中实时提取特征用于检索。范例查询允许用户从数据库中选一个用户自己认为符合查询条件的多媒体信息作为范例,也即作为用户描述的查询条件。在检索过程中,将该范例直接和数据库中的数据对象进行比较,通过这样一连串查询范例最终使用户得到需要的查询结果。例如:如果用户希望找到描述蓝天的图像,那么,他可以先在数据库中找到一个这样或类似于此的图像,查询系统将在整个数据库里搜索,给出与这个图像特征相似的查询结果。很明显,范例查询中的查询条件(也就是范例)和被查询数据应是可比较的。

范例查询在图像检索中用得很多,如果把视频数据作为图像集合来考虑而不关心它们时态上的相互关系,该方法也很容易扩展到视频检索中来。这时通常可以给出一幅图像作为范例,并将它与每个视频段的关键帧进行比较,选出关键帧符合要求的视频段作为查询结果返回给用户。与传统方法相比,这种视频查询方法有许多优点,但它仍然有不足之处:首先,这种方法仍然丢失了某些视频信息,如视频的运动信息和空间-时态特性等;另外,视频中通常包括大量的对象,每个对象由复杂的多维特征矢量来描述,这时用范例查询法来描述对象的形状和运动特性的复杂度将会很大。

范例查询是一个经典的可视化查询方法,也是一个比较成功的基于内容检索方法。许多现有的查询系统都运用了范例的方法,例如:QBIC^[5]、VisualSEEK^[6]、PhotoBook^[7]、Virage^[8]以及 FourEyes 等,但它们一般只支持静态图像查询,或对动态视频的关键帧进行查询。

(2)草图描绘法 用户通过系统提供的工具如可视化调色板,把自己需要的图像大致勾勒出来,称为草图。对于图像数据,系统直接计算草图与数据库中各图像间的相似性;对于视频检索,用户需要给出其一个具有空间特征而无时态特征的草图。用户可以描绘出查询的基本纹理、颜色、颜色的空间分布及主要对象的形状、纹理、颜色等,空间关系则由用户勾勒出的一个或多个对象之间的空间关系指定,系统计算草图和视频库中每个关键帧之间的相似性,若相似性大于某值,就认为该关键帧代表的视频段符合用户查询要求。

为了可视化描述对象的运动、空间、时态等关系,VideoQ 系统提出了一种称为“动画草图”的方法,它不仅在屏幕上画出对象的草图,同时画出这些对象的时间-空间顺序,对象的运动轨迹等,在这个方法中,运动轨迹及其持续时间、刻度也和形状、颜色、纹理等一样,成为草图中每个对象的关键属性,并作为计算索引相似性的一个条件。Bimbo 等给出另一种表达对象间空间关系的方法^[9],系统通过对象在三维空间中的空间关系,来进行二维数据的检索。他们为用户提供了在三维

空间里用数据手套(data glove)表达空间关系的方法,数据手套识别后将作为对象间空间关系并转换成二维表达。允许三维描绘更接近人类记忆方式,即使在检索二维图像数据时,该方法也是有效的。

基于草图描绘法的查询系统还包括 QBIC、VisualSEEK 等,在文[10]、[11]等中也讨论了这方面的问题。

综上所述,空间-时态描述法在传统方法的基础上对多媒体检索功能有所扩展,但显然功能有限,对于视频数据的基本可视特性,如颜色、纹理、对象形状等,都无法进行查询。而针对视频数据固有特征提出的可视化方法,能直观表达查询条件,且条件和待查询数据的特征相关,因此克服了上述缺点,检索功能更为强大,检索结果更为准确。在两种可视化方法中,范例查询更为简单,对用户的要求较少,但对视频的查询方法比较单一;与它相比,草图描绘更灵活,能表达查询条件中更细微的差别,如颜色的等级等,查询范围更广也更精确,但它要求用户能用绘画的方式基本表达自己意愿,对用户的要求更高。

5. 相关研究工作

从上面的讨论可以看出,大多数视频检索方法都是由图像检索方法扩展到视频检索中去的,这些方法为视频检索增加的功能并不多,大部分技术停留在对视频序列中的图像进行索引的阶段,而真正针对视频数据特征来进行索引的技术较少,它们更适合图像数据库的检索而不是视频的检索。下面是与视频检索相关的一些研究工作。

5.1 基于内容的视频技术

OVID 是一个面向对象的视频数据库系统,它把组成有意义场景的任意帧序列看作一个视频对象,用手工分配的属性及其取值集合来描述每个视频对象的内容。视频对象数据模型没有概要描述,为了使视频对象间可以共享描述数据,系统采用了称为“基于区间包含的继承”(interval-inclusion based inheritance)的机制。OVID 的视频对象查询包括基于帧的描述或基于区间的描述两种形式。

Little 等提出一种支持基于内容的视频检索和播放的系统。该系统定义了一个具有固定属性集的专用模式,它由电影、场景和演员关系组成,系统需要手工析取特征,并把特征插入到模式中。查询包括对电影、场景和演员属性的查询,用户可在查询结束后选择查询结果中的某部电影进行浏览。可以看出系统的查询被限定在预先定义好的属性范围内。

Nagasaka 提出一种用于自动视频索引和搜索的方法,为表示视频内容该方法在镜头内自动选取前几帧,在此基础上创建索引。但是由于最具有代表性的帧不一定总是与镜头边界的第一帧或最后一帧有关,因此还应该对视频序列进行更详细的分析以得到更好的效果。

Irain 等认为,尽管由于摄像机及物体的运动,一个视频场景中每一帧呈现出不同的画面,但它们实际上描绘了共同的场景信息,如相同的地点、相同的运动对象等,因此可以把具有相同场景信息的相关帧序列合成到一个全景图中。用户查询时先选取全景图中的静态或动态信息,而返回的检索结果则是视频序列中包含该信息内容的片段^[12]。

Hampapur 等人给出了一种对数字化视频分段的模型驱动方法以及在剪切、空间和色彩编辑时的特征析取方法。作者以视频编辑效果模型为基础,给出了标识镜头边界的形式化方法。由于在从视频流中析取有意义的信息之前我们需要识

别镜头边界,因此该工作在视频处理的初始阶段是重要的。

Manheim 大学的 Fischer 等提出了一种自动识别电影类型的方法,该方法包括三步:第一步是对颜色随机性、剪切检测、运动区域和音频随机性等特性进行语法分析;第二步是推导出型式属性,如摄像机操作的类型、场景变换和语言与音乐的表现;最后一步是将检测到的型式属性映射到电影类型上,诸如新闻片、商业片、卡通片和体育片。

5.2 多媒体信息检索系统

目前,国外已经开发了一些基于内容的多媒体/视频数据库检索系统,有的检索系统已处于实际运用当中。这些检索系统一般建立在图像基于内容的检索技术的基础上,并由此扩展了视频等方面的检索功能。

由 IBM 开发的 QBIC(Query By Image Content)系统,是一个比较成功的图像、视频数据库检索系统。它用到的检索特征包括颜色、形状、纹理、对象形状等,这些特征一般采用半自动的方法抽取出来。系统通过范例、用户描绘的草图、选择的颜色和纹理图案、摄像机移动、组成物体运动情况等方式来进行数据库查询。QBIC 将视频分成镜头,每个镜头由一个关键帧代表,视频中的运动也同时提取出来。这样,对视频特征的检索转化为对视频关键帧静态图像特征的检索,而运动仅作为图像帧的参考,并不用于对象运动的查询。

CMU 的 Informedia Project^[13],把视频处理和音频处理技术相结合,为视频图书馆建立一个智能自动索引机制。它首先使用语言处理系统 Sphinx-I 将音轨中的语言提取和识别出来,存储成全文本信息,然后通过加权技术,从文本中析取关键字。在视频处理中,系统从摄像机运动和对象出现两个方面来统计计算每个场景图像内容的重要程度,然后根据指定的压缩比例,抽取相对更重要的图像帧来代表视频内容,图像帧与相应音频信息组合起来,形成视频浏览(video skim)。一个视频浏览实际上是多媒体信息的一个简短概要。用户只需通过查找视频浏览中的关键词及交互式浏览和检索视频浏览中的图像帧,就可对信息进行查询。该技术的特点之一是采用了较多的音频识别技术作为基础,视频图像处理并不占主导地位,这和其它多媒体检索技术以视频图像作为多媒体数据处理的主要对象不同。这种技术比较适合用于教育性质的多媒体信息系统,系统在视频播放时通常同步配有解说,实际上该项目本身也就是为视频图书馆的检索而研制的。

Columbia 大学的 VisualSEEK 系统允许通过颜色及颜色区域的空间分布来查询数据库。它用颜色直方图表达整体颜色属性,根据颜色直方图的相似性度量来比较图像的颜色属性,同时该系统提供了注释图像和搜索档案的工具。

Vision 系统提供了一个有效的创建和浏览数字视频库方法。首先,用分割算法将视频分割成一定数目的逻辑意义上的视频片,接下来用音频分割处理来修正视频分割中的一些虚假检测。

Core 是一个明确支持基于内容查询的多媒体系统。它提供了一个管理复合多媒体数据的模型,一个多媒体对象由特征和概念集合来表示,同样的对象可以有多个特征,多媒体对象的检索根据特征的相似性度量集合(用来匹配特征权值),以及概念比较时的“模糊相似性”来进行。复杂查询由单个查询(属性权值、特征权值、概念权值)的加权和来完成。

此外还有斯坦福大学的数字图书馆、MIT 的 ViewStation Projects 等,目前国内尚无此类的多媒体信息检索系统^[14]。

可以看出,大部分基于内容的检索系统,其检索方法都有局限性。很多系统如 PhotoBook、Virage 和 VisualSEEk 等,都是支持静态图像的查询,而 QBIC 将动态视频的查询建立在对视频关键帧进行静态查询的基础上,其本质还是静态图像查询,而不是针对视频的动态特征的检索。其它系统也是如此,对动态视频内容的查询大部分只集中在运动对象形状、运动轨迹等运动特征这一层次的查询上。

此外,对“基于内容”中“内容”所代表的含义,各系统的理解不同,给出的解决方法也不同。有的系统将图像的物理特征(颜色、纹理等)当作“内容”;也有的把视频中对象的特征(颜色、纹理、形状、运动等)当作视频内容,尽管给出的形式很多,但并不完善,而对“基于内容”的解决都不够理想。一些重要的特征,如时态特征,在绝大多数系统里是无法体现的^[15]。即使在 VideoQ 等改进的系统中,其时态表示能力也很差,它利用用户手绘的对象运动轨迹和持续时间等特征来间接描绘时态特征,不但很不精确,而且无法明确表达各对象运动间的相对时态关系,也就无法用有效的方法把它们检索出来。更重要的是,所有这些系统几乎都完全没有涉及对视频内容的语义检索,对视频的查询集中在查询从视频表面可以看到的“现象”,不是查找这些视频“现象”所表达的意思,而这些意思往往是用户在查询过程中最需要得到的,也是用户对视频内容最直接最自然的理解。

为了克服上述系统的缺点,我们在文[16]中提出了一种新的基于内容的视频检索技术。

6. 问题与展望

多媒体基于内容检索中还有一些问题,虽然目前大部分检索方法及检索系统都还没有涉及,但我们认为随着时间和技术的发展,它们将越来越受到重视,在此列出以供探讨:

1) MPEG-4 标准 随着 MPEG-4 标准的发布,它在基于内容的检索技术方面显示了很大的潜力。MPEG-4 的编码标准是基于形状(shape)的编码,它定义了表达单个可视或声音对象的工具,以及在用户看到的最终屏幕上描述其空间-时态位置的机制。在可视对象表达上,它把传统的纹理编码与形状编码结合,描述屏幕上的任意二维对象,同时考虑了对象升级(scalability)和容错(error resilience)能力。MPEG-4 中应用的算法很大程度上类似 MPEG-2 及 H. 263 中的算法,但它在静态图像编码中还运用了小波进行编码。MPEG-4 的一个重要的新方向是它努力把自然和人工的内容集成起来,使得自然-人工混合编码成为可能。在屏幕描述上,它借鉴了虚拟现实的建模语言 VRML 的一些概念,把单个的对象联合起来组成了一个屏幕,屏幕分层描述并形成树。树结点指定屏幕结构或代表媒体对象。MPEG-4 的终端接受单个对象以及如何将这些对象在时间和空间上组织起来进行描述然后形成最终屏幕呈现给用户,而这些屏幕描述还可以动态更新。可以看出,这种描述方式对视频对象及其关系的描述很有好处,但是,如何把视频分割成语义对象仍然是一个需要探讨的问题,目前还没有一个能有效分割视频对象的技术^[17]。

2) 交互和学习 大型多媒体数据库中通常含有大量的视频图像数据,完全用手工的方式对这些数据生成索引是不可能的,检索的大部分工作,如特征的提取,对象的提取,甚至一部分的语义标注工作,都需要由计算机运用图像处理、计算机视觉等方面的技术来自动进行。显然我们很难保证计算机对每个视频流的处理结果都完全正确。因此一个多媒体检索系

统在建立后,应该通过交互的方式进行学习,减少这些错误,使检索更符合用户需求。而如何设计交互和更新操作,达到学习及减少错误的目的,还需要做进一步的研究。

3) 音频对检索的作用 前面我们把多媒体检索的重点放在研究视频的检索上。视频包含的信息量在所有媒体中最大的特点,使得视频检索一直是多媒体检索最重要的方法。音频由于本身包含的信息量大大少于视频,同时对用户来说声音较视频缺少直观性、不容易表达等因素,使得声音在多媒体检索技术中未受到人们的重视。但是我们应该借助音频分析技术,将它加入以视频为主导的多媒体检索中,利用对声音的分析结果来辅助和校正视频检索结果,以获得更好的效果。例如:场景的转换往往伴随着背景音乐曲调和节奏的变化;通过对视频流中人的语言频率的分析来确定不同的对象是否为同一个人,使得视频对象中最难匹配的对象“人”的匹配问题得到较好的解决;视频处理不确切的部分往往可以通过声音的分析加以纠正。可以看出,在视频检索中加入音频辅助检索能大大简化视频处理的复杂度,增加了视频检索的效率,是完善基于内容的视频检索技术的可行方法。

参考文献

- 1 胡晓峰,吴铃达,李国辉,等. 多媒体系统原理及应用. 北京:人民邮电出版社,1995
- 2 Chang Sh-F, et al. VideoQ: An automated content based video search system using visual cues. In: Proc. of ACM Multimedia'97. Seattle Washington USA, 1997. 313~324
- 3 Sistla A P, et al. Similarity based retrieval of pictures using indirect on spatial relationships. In: Proc. of Int'l Conf on Very Large Databases. 1994. 570~581
- 4 Allen J F. Maintaining knowledge about temporal intervals. ACM Communication, 1983, 26(11): 832~843
- 5 Flickner M, et al. Query by image and video content: the QBIC system. IEEE Computer Magazine, 1995, 28(9): 23~32
- 6 Smith J R, Chang S F. VisualSEEk: a fully automated content-based image query system. In: ACM Multimedia Conf. Boston, Nov. 1996. 87~98
- 7 Pentland A, Picard R W, Sclaroff S. Photobook: content-based manipulation of image databases. International Journal of Computer Vision, 1996, 18(3): 233~254
- 8 Hamrapur A, et al. Virage video engine. In: Proc. of SPIE on Storage and Retrieval for Image and Video Databases. San Jose, Feb 1997. 188~197
- 9 Bimbo A D, Campanai M, Nesi P. 3D visual query language for image databases. Visual Languages and Computing, 1992, 3(3): 257~271
- 10 Bimbo A D, Pala P. Visual image retrieval by elastic matching of user sketches. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997, 19(2): 121~132
- 11 Jacobs C E, Finkelstein A, Salesin D H. Fast multiresolution image querying. In: Proc. of SIGGRAPH. Los Angeles, 1995. 227~286
- 12 Irani M, Anandan P. Video indexing based on mosaic representations. Proceeding of IEEE, 1998, 86(5): 906~920
- 13 Wactlar H D, Kanade T, Stevens S M. Intelligent access to digital video: informedia project. IEEE Computer Magazine, 1996, 29(5): 46~52
- 14 钟玉琢. 多媒体信息基于特征的检索系统. 见: 第五届全国多媒体学术会议论文集. 1996. 2~6
- 15 陈慧琼, 杨敬安. 一种基于内容的视频信息检索技术的研究. 合肥工业大学学报, 2000, 23(2): 182~186
- 16 陈慧琼. 一种新的基于内容的视频/多媒体数据库检索技术的研究. 合肥: [合肥工业大学硕士论文]. 2000
- 17 Katsaggelos A K, Kondi L P. MPEG-4 and rate-distortion-based shape-coding techniques. 1998, 86(6): 1126~1132