语义视频检索综述*)

魏 维[1] 游 静[1] 刘凤玉[1] 许满武[2]

(南京理工大学计算机科学与技术系 南京 210094)[1] (南京大学计算机科学与技术系 南京 210008)[2]

摘 要 视频内容检索是多媒体应用的一个活跃研究方向,现有的内容检索技术大多是基于低层次特征的。这些非语义的低层特征难以理解,与人思维中的高层语义概念相差甚远,严重影响视频内容检索系统的易用性。低层特征和高层语义概念间的语义鸿沟很难逾越。如何跨越语义鸿沟,用语义概念检索视频内容是目前基于内容视频检索最具挑战性的研究方向。本文介绍语义视频检索出现的背景,分析语义鸿沟出现的原因,对现有尝试跨越语义鸿沟的主要方法进行综述;评述了相关技术的优缺点,探讨了各方法将来可能的研究发展方向以及视频语义检索近期、长期可能的技术突破点。

关键词 基于语义的视频检索,语义鸿沟,低层特征,语义概念,基于内容的视频检索

A Survey on Semantic-based Video Retrieval Techniques

WEI Wei[1] YOU Jing[1] LIU Feng-Yu[1] XU Man-Wu[2]

(Department of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094)[1]

(Department of Computer Science and Technology, Nanjing University, Nanjing 210008)[2]

Abstract Contend-based video retrieval(CBVR) is an active research domain in multimedia application. Most retrieval techniques on CBVR are low-lever feature based. However, these features are abstract and quite different from the semantic concepts in human thought. Video retrieval at semantic level is one of the most challenging research issues in CBVR at present. The gap between low-level features and high semantics is difficult to narrow. To go beyond low-level similarity and access video data content by semantics, we must bridge the gap between the low-level features and high-level semantics. After providing a summary about no-semantic video retrieval in the literature, this paper analyzes the reason leading to semantic gap and presents a review on the current approach to bridging the semantic gap. In addition, the future promising directions are also discussed.

Keywords Semantic-based video retrieval, Semantic gap, Low-lever features, Semantic concept, Content-based video retrieval

1 引言

近年来,数字图书馆、视频点播(Video on Demand, VoD)、远程教育和各种专业编辑系统不断广泛运用,媒体数据量呈现急剧增长的趋势[1,2]。对于含有丰富时空信息(Spatio-temporal Information)的视频数据,传统的检索技术远不能满足需要。为了能有效快捷地对海量视听信息进行过滤、浏览和检索,人们提出了基于内容的视频检索(Content-based Video Retrieval,CBVR)技术,并取得了一定的进展。CBVR已成为多媒体领域一个活跃的研究方向[3,4]。虽然人们习惯使用高层语义概念判断相似性,但是现有的视频内容检索大多是非语义层面的。由于低层特征和高层语义概念之间存在语义鸿沟(Semantic Gap),在语义概念层次进行视频内容的描述和操纵面临巨大困难[5]。如何从视频内容中提取人类思维中的语义概念,成为视频内容检索的新焦点[6]。跨越语义鸿沟,最终达到语义概念级的视频检索,正成为目前视频内容检索中最具有挑战性的研究内容[7~9]。

2 视频语义检索出现背景

当前,非语义视频内容检索的研究主要集中在相似性算法的选取和检索策略的制定,其主要类型有3种。

基于图像检索的方法:由于基于内容的图像检索(Content-based Image Retrieval,CBIR)技术相对比较成熟,所以此类方法的中心思想就是将复杂的视频内容检索转化为较容易实现的图像检索方法。把提取的关键帧(Key-frame)看作一幅图像,提取低层的图像特征后就完全采用图像内容检索的相关技术实现。此方法可充分利用现有的成熟技术,但完全未利用视频内在的显著特性——含有的丰富时空信息。为克服以上缺点,出现了第二类方法。

结合时空信息的视频检索方法:此类方法在提取低层特征的同时,注重利用能反映时空特征的信息(摄像机运动参数、视频对象的运动矢量、运动轨迹等)。比如Masahito利用了计算机视觉和图像处理中广泛应用的马赛克技术来进行视频内容的检索[3]。由于视频数据量非常大,通常视频信息都是按一定标准(MPEG, H.26x等)进行压缩后以压缩数据的

*)本研究得到国家自然科学基金(60273035)、江苏省科技攻关项目(BE2003064)资助。魏 维 博士研究生,主要从事视频内容分析、语义视频检索研究;游 静 博士研究生,主要从事视频内容分析研究;刘凤玉 教授,博士生导师,主要从事多媒体技术等研究;许满武 教授,博士,博士生导师,主要从事新型程序设计及多媒体方面的研究。

形式存在。而以上两种视频检索方法的特征提取是在像素域(Pixel Domain)中进行的,因此这些方法首先需要将视频数据解压。大量的解压计算一直是其技术瓶颈。为减少计算量,出现了基于压缩域的一些方法。

压缩域(Compressed domain)中进行特征提取的方法[10]:此类方法的思路是直接从压缩域中(或将压缩数据半解压后)提取特征信息。如 Lie 提出了在 MPEG 压缩域中物体的跟踪算法[11]。

非语义的视频内容检索查询方式主要包括两种:实例查询(Query by Example,QBE)和支持特征定义及进行草图绘制特征的查询方法(Sketch and Feature Specification)。QBE 需要访问者提供适当的视频剪辑,通过从其中提取的低层特征采用适当的相似算法进行检索。第二种方法则需要用户理解和运用抽象的低层特征。显然,这两种检索方式与人们思维中的高层语义概念存在较大差距。

3 视频语义提取方法

访问多媒体视听数据,最自然的方法是通过高层语义概念来进行操纵。语义视频检索方法的目的是利用人思维中高层语义概念(低层特征对用户不可见)来进行视频内容过滤、概要、检索。要达到语义检索的目的,关键是需要在分析和理解视频内容的基础上,用人类意识思维中的高层语义概念将视频内容表示出来,即用抽象的非几何方法表达出来[12]。

如图 1 所示,低层的特征空间包括 Visual feature、Audio feature、Text feature 等特征,这些特征一般可以自动从视频数据中提取。低层次的特征空间包含多个子空间,以 Visual feature 中颜色特征子空间为例,可以采用色度(Hue,H)、饱和度(Saturation,S)、亮度(Intensity,V)三基形成 HSV 颜色子空间。语义概念空间对应于人们通常思维中的高级语义概念。从认知层次(Cognitive Level)/角度进行视频语义划分的语义概念,主要包括事件、场景/地点和对象三类(event:火、烟、火箭发射等;scene:绿地、陆地、户外、外层空间、沙、天空等;object:飞机、船、火箭、车、鸟等)[13]。两个空间的变换不是线性变换(线性变换实质是采用一组新的基代替原空间的基),这种变换/映射很难用数学方法描述并建立模型,这是语义鸿沟出现的根本原因。尽管计算技术不断发展,但让计算机准确地理解视频中的语义概念仍是个难题,因此要完全跨越语义鸿沟是十分困难的。目前提取视频语义的主要方法包括概率(Probabilistic)统计方法、统计学习(Statistical Learning)方法、基于规则(Rule-based)推理的方法、结合特定领域(Domain-dependent)特点的方法等。基于语义视频检索系统的实现主要包括两大类技术:视频语义内容分析、语义提取。

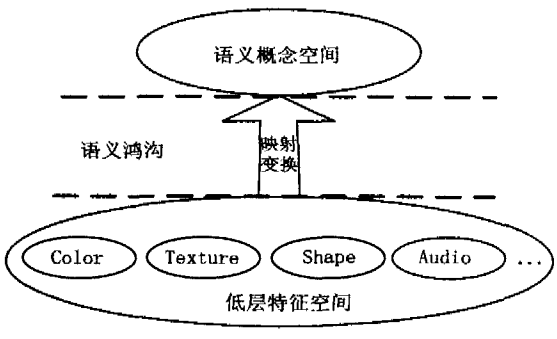

图 1 语义鸿沟

3.1 视频语义分析

视频语义分析包括时域分割(分镜头)、空域分割(区域分割、视频对象分割)、关键帧选取及语义特征的提取等。

视频镜头(Video Shot)是在摄像机一个拍摄动作(Camera Action)中所录制连续视频帧所组成的视频序列。在这些帧序列中,彼此间内容有很强的相关性,因此通常将镜头作为视频内容组织的基本单位。目前分镜头的技术研究已经比较成熟,镜头边界的探测算法的效率和准确度也不断提高。镜头边界分为突变和渐变两种,其中突变类边界占大多数(90%以上)。渐变主要是由于溶解(Dissolving)、淡入\出(Fading In\Out)等特殊技术处理形成的。对于突变类的边界检测,在像素域和压缩域中的效果都比较满意。渐变类的边界检测算法在像素域中有一定突破,然而在压缩域中的算法还有待提高[14]。Taskiran 等在系统 ViBE 中直接从压缩域的 DC 序列中提取 GT(Generalized Trace)和衰退树(Regression Trees)进行镜头分割[15]。Ouyang 在 MPEG 压缩域中用 MBs 和运动矢量(Motion Vector,MVs)等信息进行体育比赛中回放镜头的边界探测[16]。以上两种方法都是在压缩域中进行的,而 Wengang 充分利用视频中声音在镜头边界有改变的特征,同时联合音频和可视图像进行边界探测,提高了镜头分割的准确性[17]。

为了从媒体数据中得到语义概念,往往需要将空域分割的视频内容用特征量的形式表示。这些特征的提取可降低语义模式匹配和识别的难度。语义对象往往与帧中所对应区域的特征有很强的关联,所以对于语义的提取而言,区域分割和视频对象分割具有重要作用[18]。抽取的特征最好是在模式类间具有不变的性质。只有提取显著区分特性的特征才可能用简单模式分类法进行分类和识别。区域分割和视频对象分割对于语义概念抽取具有明显区分意义,是视听信息在时空上分割比较重要的环节。

在区域分割方面,现有技术可对均质区域实现自动分割、并且分割准确度较高[19]。Rautiainen 在局部 HSV 颜色直方图的基础上,建立自组织图(Self-organizing Map),通过对其训练来探测皮肤的区域[20]。Sigal 等提出一种在视频序列中实时分割皮肤区域的新方法[21]。然而,对语义概念提取有较大作用的区域往往是复合颜色并且是非均质的,其实现需要人干预调整参数。

视频对象分割方面,现阶段可实现用户监督下的半自动视频对象提取[22]。例如,Chen 在视频对象分隔时采用首帧分割,自动对象跟踪和边界精化技术,在用户监督下此方法可实现高效率的半自动视频语义对象分隔[23]。虽然语义提取中急需的快速自动进行视频对象提取当前还不成熟,但此方面的探索性研究也比较多。Kai 提取光流场的相直方图,以此探测视频帧中的语义对象,最终实现非监督的视频语义对象提取[24]。Zhou 通过区域提取和运动预测两项技术,解决了视频语义对象提取的速度问题[25]。Lievin 等在联合处理低层颜色和运动量的基础上,进行颜色空间的非线性变化[26]。

低层特征一般用特征描述子表示。描述子(Descriptor)是刻划特征的一个数据结构,一个描述子的维数可以是多维的。常用的可视特征描述子包括:颜色描述、纹理描述、形状描述、运动描述[27]。视频特征大都是高维(如 MPEG-7 中颜色结构描述子达 128 维)。为准确提取语义概念、减少计算量和避免维数灾难(Curse of Dimensionality),常要做降维处理。降低维数可以采用特征提取和特征选择两类方法。特征提取

通过映射(变换)的方法用低维空间表示样本。而特征选择则从一组特征中挑选最有效的特征,以此达到降低特征空间维数的目的[28]。通常特征子集的产生方法(策略)是穷举法和启发式方法[29]。穷举法把各种可能的特征组合都算出来,通过比较选择最优的特征组,其典型的算法是 FOCUS[30]。启发式的选择方法依据特定的启发策略来增减搜索空间[31]。

特征提取、选择虽然在理论研究上已经比较成熟,但新的实现方法近年来不断出现[32]。Balaji 选择与分类最相关的联合特征,提出了在高维数据分类中性能优良的联合分类特征最优算法(Joint Classifier and Feature Optimization, JCFO)[33]。然而 JCFO 算法计算复杂,并不适合视频数据应用。其他现有的特征提取、选择算法往往也不能直接用于视频,因此有必要结合视频特征的性质进行降维研究[34]。对于视频中高维描述子,Ankush 提出了基于最小分类错误的 DABER 算法,同时提出了减少描述子的 CPDDR 算法[35]。文中的思想对今后视频领域的降维研究有一定启发作用,但 Ankush 在上文中并未进行涉及语义保持的研究。而在文[36]中,Jensen 和 Shen 用基于 rough 集的方法,对数据进行语义保持(Semantics-preserving)的降维处理,此文为视频数据保持语义的降维研究提供了可借鉴的方法。

3.2 概率统计方法

概率统计方法将视频语义对象提取看作是待提取视频语义对象(此对象类别未知)的分类问题,利用模式分类方法来尝试跨越语义鸿沟。

语义检索的随机方法关注的是模型概率特性,其核心思想是用随机数学方法来描述对象的不同特征并在此基础上建立多媒体概念模式分类器。如图 2 所示,视频语义概念模式的分类器主要包括多媒体语义对象模型和多媒体语义网络模型。建立分类器的过程要涉及两方面,即给定一般的模型或分类器的形式及利用训练样本去学习或估计模型的未知参数[37]。

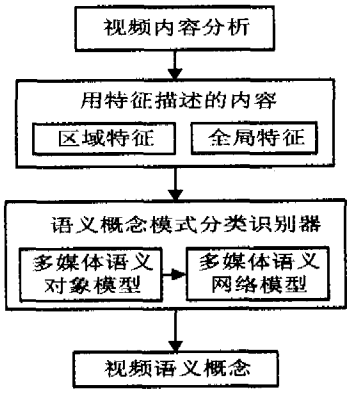

图 2 语义视频检索随机方法框图

3.2.1 多媒体语义对象模型

视频的任意部分(片段)内容都可以理解为在某一地点或场景下存在或发生的事件。依据此理解,提出多媒体对象的语义概念。多媒体对象是多种层次特征(即包括低层次的声音、图像、字幕特征,也包括分割的特征,还包括诸如人脸识别器等高层次的特征探测器)所支持的一种概率模式[38]。多媒体对象利用概率结构模型作为中介,使低层次特征和高层次语义概念间产生联系。通常来讲,多媒体对象不仅在帧内的空域具有空间上的随机性,而且在每一个帧的时间序列及音频时间序列中还具有时间、空间上的随机性质,所以通常在此模型中将低层的特征作为一个随机变量 X(矢量)。一般可以采用贝叶斯决策理论(Bayesian Decision Theory)建立的贝叶斯分类器来作为语义对象分类模型[39,40]。

具体来讲,可以把观察到的特征值表示为多维随机变量 X(向量),定义可能的假设 H(较简单的方法可以采用定义两个假设 H_0 和 H_1,其中 H_0 表示语义概念对象出现,H_1 表示语义对象未出现)。对每一个假设,定义特征的条件概率密度函数和先验概率。通常用贝叶斯决策理论在可能的假设间做决策时,认为条件概率密度函数是已知的[35]。对静态地点类语义概念,可用高斯混合模型(Gaussian Mixture Models, GMMs)来得到其条件概率密度函数。对于同时具有时空关系特性的事件和对象而言,用隐马尔可夫模型(Hidden Markov model, HMM)得到每种假设下对应的条件概率密度函数。

由于隐马尔可夫模型在语音识别方面应用效果较好,所以目前主要采用 HMM 建立多媒体声音对象、事件模型。在视频中的声音往往是多个不同声音源的合成(比如背景音乐和前景声音往往同时存在),混合音源中提取语义概念是音频语义的主要研究内容。

3.2.2 多媒体语义网络模型

用语义对象模型分类得到的语义概念之间并不是相互孤立的。在视频内容的上下文背景中,语义概念间存在彼此的联系。多媒体对象网络就是描述对象间这种强关联性。比如,天空、雪出现在户外的概率较大,人讲话时往往伴随嘴唇的活动等。为描述帧层次上语义概念间的关系,可用加权图(Factor Graphs)来建立模型[41,42]。加权图包括贝叶斯置信/信念网(Bayesian Belief Network, BN)和马尔可夫随机场(Markov Random Field),其中用得较多的是贝叶斯置信网。BN 是描述联合概率分布的有向无环图(Directed Acyclic Graph, DAG)的拓扑形式。贝叶斯网络是用来表示变量间连接概率的图形模式,它提供了一种自然的表示因果信息的方法,用来发现数据间的潜在关系。在这个网络中,用节点表示语义概念,有向边表示语义概念间的依赖关系[43,44]。

语义网络可以间接提高模式分类器的语义概念识别能力,一些难以直接探测的语义概念可以通过其他容易探测的相关对象推理而得[45,46]。比如海滩概念难以直接探测得到,但海滩的景色常伴随水、沙、树、船等容易识别的语义对象。因此可以通过水、沙、树、船间接得到海滩语义,同时推断出这是一个户外景。

网络模型目前应用较成功。Hoogs 在处理视频中大量的对象、事件和场景时,将语义对象分类与语义数据库相结合。由于此方法结合电子词典数据库 WordNet,所以对象和事件的识别能力得到大大增强[47,48]。Cheng 提出基于语义网络的语义联合模型[43],此模型利用不同镜头间对象的关系来描述镜头间内容的相互关系,在形式上用六元组定义的联合模型表示。Luo 用动态贝叶斯网(Dynamic Bayesian Network, DBN)和层次隐马尔可夫模型(Hierarchy Hidden Malkov Model, HMM)建立由粗到精的语义概念模型[49]。Wang 等在文[50]中以智能代理(intelligent agents)确定网球球的轨迹和落点,以之作为改进的贝叶斯网络分类特征,分类后得到语义标签。

3.2.3 模式分类器的训练和学习

利用样本数据来确定分类器的过程称为训练分类器。在多媒体语义对象模型方法中需要用到 EM 算法来估计期望、协方差矩阵、GMM 和 HMM 的混合比例(Mixing Proportions),以及 HMM 中的转移矩阵。在机器学习算法的经验分析方面,可利用 UCI 机器学习知识库中的数据[51,52]。

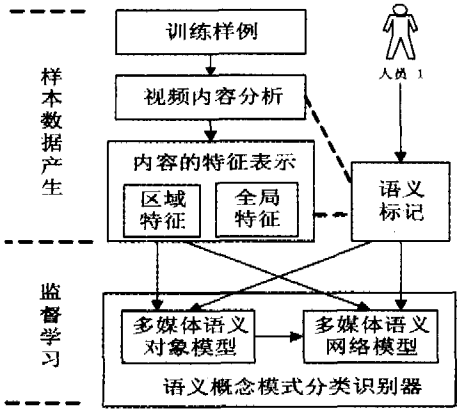

图 3 语义概念模式分类器学习过程

语义概念模式分类器学习过程如图 3 所示。待学习的语义概念或函数称为目标概念(Training Concept),记作 c。一般来说 c 可以是定义在实例集上的任意布尔函数,即 c: $X \to \{0,1\}$。概念定义在一个实例(Instance)集合之上,这个集合表示为 X。在学习目标概念时,必须提供一套训练样例(Training Example),每一个样例为 X 中的一个观察值 x 及其目标概念值 $c(x)$。对于 $c(x)=0$ 的实例称为反例(Negative Example)或称为目标概念的成员。对于 $c(x)=1$ 的实例称为正例(Positive Example)或称为非目标概念的成员。训练样本集中每个样本的类别归属是(在人的参与下)"被标记了"的(Labeled),通常在语义训练中用到的是有监督(Supervised)学习。分类器学习的目标就是寻找一个假设 h,使对于 X 中的所有 x,$h(x)=c(x)$[53,54]。

3.3 统计学习方法

语义概率方法(传统的统计模式识别方法)研究的是样本数趋向无穷大时的极限特性,是一种渐进理论。其性能在样本数足够多的前提下才能达到理论效果。而视频检索中样本数目往往有限,因此如何应用有限样本情况下的统计学习理论进行语义概念提取,也是研究的重点之一。

支持向量机(Support Vector Machine,SVM)基于统计学习理论,建立在计算学习理论的结构风险最小化原则之上。其目的是在高维空间中寻找一个超平面作为两类的分割,以保证最小的分类错误率。此类模型在只有小训练样例集的情况下,分类效果较好[13]。如 Naphade 利用 SVM 作为主动标注和主动学习的内在分类器[55]。这种以支持向量机为基础的标注器建立在少量已标注的数据之上,每次新数据学习后,分类器参数都会相应更新。

3.4 基于规则推理的方法

以上两种方法的理论基础都是模式分类,实质上是分类器通过学习训练样例由系统内部产生分类标准。而基于规则推理的方法则考虑直接从系统外给定分类标准,即规则。

基于规则推理方法可以定义为集合 R。

R: $F \to C$(F 是特征集合,C 是语义概念集合)。

若对于 $f \in F$ and $c \in C$,

c 依赖于 f

则存在一个规则:$f \to c \in R$。

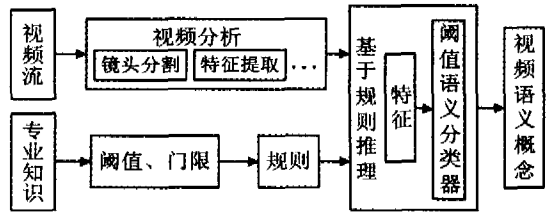

图 4 基于规则的语义提取方法

图 4 为基于规则的推理方法主要组成框图。此类方法主要是根据视频内容特点,结合专业知识(往往由专家参与)定出相关的推理规则。推理规则的实质是给出语义分类的阈值(Threshold),利用一系列的门限值构成语义概念分类器。不同内容的视频流经视频分析后以镜头为单位提取特征,进行推理分类并提取对应的语义概念。语义事件规则的制定有两类:一是依据可视特征和时空关系来定制,另一种是根据对象在现实中的关系(Real-world Relation)定制。同时再加上时域上定义的相互位置关系和逻辑运算关系,便可实现预定义的语义对象、事件的检索。确定性事件选择首先以低层的视听线索/特征来表示事件,比如足球比赛中采用场地颜色、摄像机运动和边界等表示。在这些具有特征性质的线索被探测后,再根据针对特定领域所制定的规则进行推理,得出语义概念。

传统的规则方法一般都遵从图 4 所示的方法进行语义提取。Petkovic 定义特征操作符(f)、空间关系(s)、时间关系(t),在此基础上定义对象规则(Object rule)[56]。Tiecheng 在教学视频中定义局部改变、内容一致和重要改变规则,进行语义内容概要[57]。这些方法通常采用单规则提取语义,其语义概念单一。而在传统规则方法基础上改进提出的多规则语义提取方法,可支持较丰富的语义概念。如 Li 在对美式足球训练视频分析时,利用帧内绿色像素的比例、镜头时间的长短、是否包含动作等线索制定向前推理规则,设立多门限值,实现语义事件识别[5]。

实际系统的现实过程中,一个规则的准确制定十分困难,近两年模糊理论逐渐用到语义规则的抽取中。Dorado 等利用模糊集理论和数据挖掘技术,实现了基于规则的语义提取和标注[58]。学习阶段,此方法在专家用户参与下定制小量的样本,以此建立规则和知识库。利用规则库,系统自动提取和标注视频中的语义概念。Ho 结合智能遗传算法,提出紧凑的模糊规则分类系统,其模糊规则可理解性好、系统准确度高[59]。将模糊理论应用在规则的制定中,是今后规则推理方法中一个具有良好发展前景的研究方向。

3.5 结合特定领域特点的方法

通过限定、缩小视频领域(Narrowing the Domain)是目前跨越语义鸿沟的有效方法之一。限定特定的领域后,语义概念和事件的随机性就被缩小了[60]。以上是此方法的理论基础。结合特定领域特点的方法建立在对特定领域视频数据独立性质的分析上。通常结合应用领域的背景知识,简化从低层特征到高层语义概念的映射关系。具体方法大多利用视频对象的位置、对象在时间轴上的变迁,与特定语义事件的关联等来实现。

此类方法提出时间较早,目前应用效果也比较满意。

Rautiainen 利用建筑对象在垂直和水平方向的直方图累计量较大的特性，用边界方向相关矢量(Edge Direction Coherence Vector)或统计的边概率直方图(Edge Direction Histograms)来区分自然景色和城市语义[61]。在外科视频教学领域，Luo 通过混合高斯模型进行语义原则视频镜头分类，用混合高斯模型近似模拟数据的分布，用自适应的最大期望(EM)算法选取高斯参数[62]。在新闻节目中，Fan 建立层次语义概念分类器，每个节点对应一个语义概念[63]。在影片语义分析领域，Rasheed 等结合影片的特点只用四个视觉特征(平均镜头长度、颜色差异、运动内容和灯光)将电影分为悲剧、动作、戏剧和恐怖片几种类型，达到影片语义分类的目的[64]。

3.6 其他方法

其他方法包括基于注释(Annotation-based)的方法、语义模板方法、多种语义线索结合的方法等[65]。这些方法对视频语义检索的发展和探索来讲具有很好的借鉴作用和启发性。

音频语义和视频语义的联合应用方面也出现了不少新方法：Asano 用自适应波束形成(Adaptive Beamforming)将声音按音源位置分类，然后结合可视特征探测视频中讲话的语义事件，这是综合声音与可视信息进行语义分类的典型应用[66]。Miyauchi 综合利用时域上语义相关的字幕、声音和可视特征进行语义事件探测[67]。

Gillespie 和 Nguyen 提出用活动能量流(Activity Power Flow)来粗略描述视频镜头的空间内容及时域上内容的变化[68]，同时提出直接从 MPEG 压缩数据中计算运动密度的改进算法。此方法最终可高效地把镜头内容分为体育、戏剧、景物、新闻等高层语义概念。Ekin 探索用图的结构来建立、描述和表示语义概念及语义事件[69]。Wang 等在系统中引入中间件模块，实现语音在语义检索中的应用[70]。

上述语义概念大多是从认知层次/角度进行划分和提取的，研究主要集中于提高语义分析和语义提取的有效性和可靠性。而近年来也出现了从情感层次(Affective Level)进行情感语义提取的研究。情感语义与人的心理感受关系密切，描述人心理感受的视频语义(如浪漫风景)。Juhani 把情感语义也纳入视频语义范畴，结合媒体组语言情绪集(一个关于情感的数据库)自动识别口语中的情感语义[71]。与 Juhani 的方法不同，在文[72]中，Hanjalic 等则以心理学中情感维数为基础提出情感视频语义的表示和建模计算框架，把视频情感语义表示为二维(Arousal，Valence)情感空间上的点集。用连接 arousal、valence 与视频低层特征的模型将情感语义映射到二维空间，实现了视频情感内容的提取。

其他的一些新方法也在视频语义概念探测中得到应用：Osadchy 把新近提出的 anti-faces 探测方法进行延拓，用在视频事件探测上，把时间序列上的变化看作是三维的仿射变换(Affine transformation)：$f(x,y,t)\rightarrow f(x,y,at)$[73]，以此进行视频事件识别。Ekin 等则在传统 ER 模型基础上进行语义概念扩充，建立语义句法模型，并通过基于图的检索模板进行语义检索[74]。

4 多模式融合、多层次语义分析

视频综合应用图像、声音、文字等信息表达特定的主题。采用多模式融合(Multimodal Fusion)和多层次分析(Multilayer Analysis)技术进行视频语义提取，是尝试跨越语义鸿沟的有效途径之一[75]。

现有的多模式融合技术分为特征融合/早期融合(feature fusion)和决策融合/晚期融合(decision fusion)。将场景镜头中提取的多模式特征(图像、声音、文字等)作为后续语义提取模型的输入，即特征融合。这种融合往往产生高维特征矢量，通常需降维处理。语义概念所涉及的多模式特征间的关联程度各有不同，所以这种融合方法对语义提取的效果并不理想。决策融合却采取在较高层次集成各种语义提取的模型(方法)的策略，产生一个整体/全局的语义决策。在这类方法中主要采用机器学习方法进行融合较多。

直接将低层特征映射到高层语义概念很难实现，而多层次分析可将高层语义分解为一系列可识别的低层原型(Primitives)及各原型和高层语义的约束关系。低层基本事物与低层特征可直接产生映射。从低层原型事物出发，通过推断便可提取语义概念。例如，Fan 等在文[76]中采用多层分析建立中间层的语义视频概念，用显著对象(Salient Objects)作为低层特征与高层语义间的中间层，在医学视频中推理提取语义概念。现有文献中多层次分析大多采用统计方法，其中采用 HMM 等概率图模型进行推理较普遍。

5 评测标准与相关压缩标准

为不断推动语义视频检索的深入研究，需要对各种语义提取模型性能进行横向对比。TREC (Text REtrieval Conference)一直致力于为大规模的信息检索提供有效的评测方法。2001 年起，TREC 对基于内容的视频检索提供大量的测试集，同时为 CBVR 性能的评估提供统一的标准。视频检索评测(TRECVID)从 2002 年起开始建立并完善了语义概念评测的相关标准，对高层的语义探测方法性能进行评测。TRECVID 为语义视频检索提供开放统一的评测标准，方便了各系统间的性能比较。每年的评测指导方针(Guidelines)对语义视频检索的研究内容、方向，特别是如何评测具有相当的影响[77]。

广泛采纳的视频标准中也逐渐提供对语义的支持。如 MPEG-4 在压缩编码中引入了视频对象[78]，多媒体内容描述接口 MPEG-7 中较全面扩充了支持语义的低层特征描述子[27]。这些标准的制定有利于视频语义检索系统的实现。

6 讨论

以上主要介绍了各种语义提取方法实现的相关技术。生物的语义模式分类系统，构造精密、功能完美，但要想在视频检索中完全达到人类思维中的语义概念水平是不可能的。目前的各种方法在不同侧面、不同程度上尝试跨越语义鸿沟的限制，每种方法各具优点和不足之处。这些探索性方法为语义视频检索的深入研究奠定了一定基础。

6.1 语义分析

特征提取：特征提取通常是要提取对语义分类器或语义抽取最具有鉴别(Distinguishing)能力的特征。这种鉴别能力的标准是：来自同一类别的不同样本的特征值应该非常相近，而来自不同类别的样本的特征应该有很大差异。这些特征对于类别信息不相关的变换具有不变性(Invariant)。通常颜色直方图是图像检索领域广泛应用的可靠特征，其对尺寸、形状、帧内运动等都具有良好的不变性，并且对噪声的敏感性也较低。

空域分割：语义对象在帧中都占据一定的空间，这些空间区域对应的某些特征量具有良好的语义鉴别性质。因此，区域分割和视频对象分割对于语义视频检索来讲具有重要意

义。

空域分割与视频对象分割关系密切，空域分割技术的发展制约着视频对象分割技术的发展。基于单一特征（颜色、纹理等）进行的均质区域分割技术必然只能适用于简单的视频对象提取。今后一段时间，联合利用颜色、纹理和运动等特征信息进行多颜色、多运动的前景/背景分割，仍是视频分析中一个有价值的研究方向。

视频数据降维：媒体内容是高维的数据，降维不仅可以减少计算量，而且可提高分类精度。视频中常采用选择特征方法，即选择特征集中较少的元素组成特征子集，用此子集组成的空间作为新的描述空间。在特征变换方法中，一般常用的是主成分分析（Principal Component Analysis，PCA）方法。此方法建立以原特征集作为定义域的线性或非线性函数。因为这些函数的值域比原特征集的元素少，所以可间接降低维数。但从分类的角度来看，PCA 变换得到的特征并不适合语义概念提取。因此在语义视频中只能用其它的线性或非线性变换降维。

保持视频语义与数据进行维处理在某些情况下是相互矛盾的。如何兼顾二者，在保持语义的前提下结合视频的特点进行有效降维，是目前视频降维研究急需解决的难题。

6.2 各方法评述

概率统计方法：此方法的出发点基于统计概率，以其建立的语义模型对媒体中大量的随机性内容加以描述有较好的效果。多媒体语义对象分类模型的目的就是要将一个语义概念（由它的特征表示）判别为它所属的某一类。采用贝叶斯分类器时，语义概念是按最大后验概率进行分类的，这由一个判别函数来完成。多数情况下，该判别函数是线性的或二次的。当类服从正态分布时，要找到最优线性分类器总是不可能的。就目前所知，都是协方差矩阵相等的情况。

随机模型中加入学习/识别模块，主要是为了能反映媒体内容本质的非确定性。目前，学习方法以监督学习为主。在将来的语义视频检索系统中，为达到简洁表示语义概念的目标，还需充分利用监督学习、非监督学习、半监督学习、强化学习（Reinforcement Learning）方法来进行训练。

在今后的随机方法中，为提高系统的语义概念识别能力，应充分利用视频中多种语义线索（声音、字幕、可视图像等）来建立语义分类器。

统计学习方法：支持向量机方法得到的分类器，其重要优点是可以采用支持向量的个数而不是变换空间的维数来表征。SVM 只需要较少的训练样例就可以估计和决定分类器，并且较少产生过学习现像。但是有限的小训练样例集对于高维的视频特征空间而言，几乎趋近于空集。因此，SVM 虽然在训练时的性质很好，但在学习后用于分类时的效果就大打折扣了。

基于规则推理方法：此方法相对于随机模型方法而言，实现较容易、计算效率高。在一定程度上可跨越语义鸿沟的限制，能在基于内容的视频检索系统中实现语义层次的操纵。由于阈值类型是预先定义好的，因此语义概念的种类固定，难以满意地描述视频内容中大量随机出现的语义概念。一般来讲，定义语义对象的规则相对容易，但要给定复杂语义事件的规则就相当复杂（尤其在基于多线索/特征的推理准则制定时显得更突出），这种复杂事件的分类器的效果也并不乐观。在实施中往往将固定值作为门限，而实际的真实情况是门限值随着视频内容的不同有一定的变化，因此定值将在“源头上”带来误判断。实际系统中，如何选择最优的门限阈值是很难解决的问题。

今后，规则推理方法可充分借鉴和模拟人的规则推理过程。应用模糊逻辑、规则挖掘技术并结合自主学习方法，实现动态创建低层特征和高层语义间的联系规则，使得规则集中元素可以实现自动更新，突破预先定义语义规则数量的限制。同时，如何解决在多规则推理时规则间的重叠和冲突问题也待深入研究。

结合特定领域特点的方法：此方法实现了直接从低层次特征到高层次语义概念的跨越，并且实现较容易，语义映射效果也令人满意。然而，此类语义映射（方法无通用性）只适应用特定视频节目，而且对一些复合性的事件难以识别。将来，此方法的应用主要集中在内容随机性变化较小的节目中，如新闻、体育类节目。在结合特定领域特征简化映射关系上，此类方法还存在较大的探索空间。

结束语与未来研究方向 本文介绍了语义视频检索技术的发展动态及研究内容和方法。目前视频检索中用的语义概念划分主要局限于认知层次，而在今后的研究中，可考虑情感等语义的结合，同时利用认识模型、文化背景、美学标准、冷暖色调来表达一定的感情[48]。这些语义的不断丰富，有利于视频内容的语义细化描述。

现有的语义概念提取方法各有特点，其目的是为了建立较好的预测数据模型。多类特征（可视特征、声音特征、时间线特征等）的融合，多种方法（概率方法、统计学习方法等）得到的不同语义概念在同一检索系统中进行融合，是今后研究的两个发展方向。如何建立多层次融合的体系结构，得出更一般的视频语义，是近期有发展潜力的研究方向。各种方法的相互交叉应用，也是以后缩小语义鸿沟的一个有效途径。

从长期来讲，进行仿生研究可能使语义视频检索得到突破性进展。事实上，从生理认知实验中发现，人的概念分类和对象识别并不是在特征空间上定义不同对象的分类，也不是符号化表示的数据理解方法，而是从记忆中的经历（hand-on）的实例中抽象生成原形模式。这种直觉理解源于体验并伴随实例模式的记忆，同时还有大量的相似性比较[79,80]。如何仿照和模拟人的意识分类过程、机制（并产生不同类别决策边界），探索新的跨越语义鸿沟的方法，将是语义视频检索在较长时期中的研究方向。

低层特征与高层语义概念间的变换/映射关系的研究现在还处于起步阶段，各种不断涌现的方法具有尝试性质。将来，随着变换/映射关系的深入研究，相信视频语义检索方法会不断走向成熟，并最终会给基于内容的视频检索带来深刻的技术变革。

参 考 文 献

1 Chen C-S, Hsieh W-T, Chen J-H. Panoramic Appearance-Based Recognition of Video Contents Using Matching Graphs. IEEE Transactions on Systems, Man and Cybernetics, 2004, 34(1): 179～199

2 Peng Y X, et al. Video clip retrieval by maximal matching and optimal matching in graph theory. In: 2003 International Conference on Multimedia and Expo, Baltimore USA, 2003. 317～320

3 Hirakawa M, Uchida K, Yoshitaka A. Content-based video retrieval using mosaic images. In: IEEE Proceedings of First International Symposium on Cyber Worlds (CW'02), Tokyo, Japan, 2002. 161～167

4 Assfalg J, et al. Semantic annotation of sports videos. Multimedia, IEEE, 2002, 9(2): 52～60

5 Li B, Sezan I. Semantic sports video analysis: approaches and new applications. In: IEEE Proceedings of 2003 International Confer-

ence on Image Processing, Barcelona, Spain, 2003. 17～20
6 Wu C, et al. Events recognition by semantic inference for sports video. In: IEEE Proceedings of First International Symposium on Cyber Worlds (CW'02), Tokyo, Japan, 2002. 805～808
7 Babaguchi N, Nitto N. Intermodal collaboration: a strategy for semantic content analysis for broadcasted sports video. In: IEEE Proceedings of 2003 International Conference on Image Processing, Barcelona, Spain, 2003. 13～16
8 Adams W H, et al. Semantic indexing of multimedia content using visual, audio and text cues. EURASIP J Appl Signal Processing, 2003, 1(1): 170～185
9 Lu S, Lyu M R, King I. Semantic Video Summarization Using Mutual Reinforcement Principle and Shot Arrangement Patterns. In: Proceedings of the 11th International Multimedia Modelling Conference, 2004. MMM 2005, Melbourne, Australia, 2005. 60～67
10 Jiang J, Weng Y. Video extraction for fast content access to MPEG compressed videos. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(5): 595～605
11 Lie W-N, Hsiao W-C. Content-based video retrieval based on object motion trajectory. In: 2002 IEEE Workshop on Multimedia Signal Processing, Virgin Islands, USA, 2002. 237～240
12 Liu J, Bhanu B. Learning semantic visual concepts from video. In: IEEE Proceedings of 16th International Conference on Pattern Recognition, Quebec, canada, 2002. 1061～1064
13 Lin C-Y, Tseng B L. Segmentation, classification and watermarking for image/video semantic authentication. In: 2002 IEEE Workshop on Multimedia Signal Processing, St Thomas, Virgin Islands, USA, 2002. 359～362
14 Mezaris V, et al. Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval. EEE Transactions on Circuits and Systems for Video Technology, 2004, 14(5): 606～621
15 Taskiran C, et al. ViBE: A Compressed Video Database Structured for Active Browsing and Search. IEEE Transactions on Multimedia, 2004, 6(1): 103～118
16 Ouyang J-q, Li J-t, Zhang Y-d. Replay boundary detection in mpeg compressed video. In: 2003 International Conference on Machine Learning and Cybernetics, Washington, DC, USA, 2003. 2800～2804
17 Wengang C, De X. Content-based video retrieval using audio and visual clues. In: 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering, Beijing, China, 2002. 586～589
18 Park Y. A framework for description, sharing and retrieval of semantic visual information: [Ph D dissertation]. Tucson, 2002
19 Ekin A, Tekalp A M. Robust dominant color region detection and color-based applications for sports video. In: 2003 International Conference on Image Processing, Barcelona, Spain, 2003. 21～24
20 Rautiainen M, Seppanen T, J P J P. Detecting semantic concepts from video using temporal gradients and audio classification. In: International Conference on Image and Video Retrieval, Urbana, IL, 2003. 260～270
21 Sigal L, Sclaroff S, Athitsos V. Skin Color-Based Video Segmentation under Time-Varying Illumination. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(7): 862～877
22 Sun S, Haynor D R, Kim Y. Semiautomatic video object segmentation using VSnakes. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(1): 75～82
23 Chen H, Qi F, Zhang S. Supervised video object segmentation using a small number of interactions. In: 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, China, 2003. III365～III343
24 Ma K-K, Wang H-Y. Unsupervised semantic video objects segmentation over optical-flow field. In: 7th International Conference on Control, Automation, Robotics and Vision, Singapore, 2002. 1216～1221
25 Zhou K, et al. Fast tracking of semantic video object based on motion prediction and subregion extraction. In: IEEE Proceedings of 2002 International Conference on Image Processing, Benalmadena, Malaga, Spain, 2002. 621～624
26 Lievin M, Luthon F. Nonlinear color space and spatiotemporal MRF for hierarchical segmentation of face features in video. In: IEEE Transactions on Image Processing, Santa Clara, California, 2004. 63～71
27 JTC1/SC29/WG11 I. L. Coding of Moving Pictures and Audio, "Overview of the MPEG-7 Standard" Int'l Organization for Standariation. 2000
28 边肇祺, 张学工, 等. 模式识别. 北京: 清华大学出版社 2000
29 Zhao J, et al. The study on technologies for feature selection. In: 2002 International Conference on Machine Learning and Cybernetics, Beijing, 2002. 689～693
30 H A, T G D. Learning Boolean concepts in the presence of many irrelevant feature. Artificial intelligence, 1994, 69(1): 278～305
31 A L M, Langley. Selection of relevant features and examples in machine learning. Artificial intelligence, 1997, 97(1): 245～271
32 Lynch R S, Willett J P K. Classification performance results of Various medical diagnostic data sets. In: Component and system diagnostics, prognostics, and Health Management, 2002. 80～87
33 Krishnapuram B, et al. A bayesian approach to joint feature selection and classifier design. IEEE Transactions on Systems, Man and Cybernetics, Part B, 2003, 33(3): 448～464
34 Mittal A A, Cheng L -f. Achieving semantic coupling in the domain of high-dimensional video indexing application. In: SPIE Conf. Applications of artificial neural networks in image processing, 2001. 97～107
35 Mittal A, Cheong L-F. Addressing the problems of Bayesian network classification of video using high-dimensional features. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(2): 230～244
36 Jensen R, Shen Q. Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(12): 1457～1471
37 Duda O, 等. 模式分类(第二版). 北京: 机器工业出版社, 2003
38 Naphade M R. A probablistic framework for mapping audio-visual features to high-level semantics in terms of concepts and context: [Ph D dissertation]. Computer Science, Urbana-Champaign, 2001
39 Vailaya A. Semantic classification in image databases: [Ph D dissertation]. Lansing, 2000
40 Thomaz C E, Gillies D F, Feitosa R Q. A New Covariance Estimate for Bayesian Classifiers in Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(2): 214～223
41 Naphide H R, Huang T S. A probabilistic framework for semantic video indexing, filtering, and retrieval. IEEE Transactions on Multimedia, special issue on Multimedia, 2001, 3(1): 141～151
42 Ramesh N M, Kozintsev I V, Huang T S. A factor graph framework for semantic video indexing. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(1): 40～52
43 Cheng Y, Xu D. Content-based semantic associative video model. In: 6th International Conference on Signal Processing, Beijing, China, 2002. 727～730
44 Zhou L, et al. Applying the Naive Bayes Classifier to Assist Users in Detecting Speech Recognition Errors. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences, Hawaii, America, 2005. 183b～183b
45 Shih H-C, Huang C-L. Image analysis and interpretation for semantics categorization in baseball video. In: Proceedings of International Conference on Information Technology, Las Vegas, NV, USA, 2003. 379～383
46 Greenspan H, Goldberger J. Probabilistic space-time video modeling via piecewise GMM. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(3): 384～396
47 Hoogs A, et al. Video content annotation using visual analysis and a large semantic knowledgebase. In: Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA, 2003. II-327～II-334
48 Stein G C, Rittscher J, Hoogs A. Enabling video annotation using a semantic database extended with visual knowledge. In: IEEE Proceedings of 2003 International Conference on Multimedia and Expo, Washington DC, USA, 2003. 161～164
49 Luo Y, Hwang J-N. Video sequence modeling by dynamic bayesian networs: a systematic approach from coarse-to-fine grains. In: 2003 International Conference on Image Processing, 2003. 615～618
50 Wang J R, Parameswaran N. Analyzing Tennis Tactics from Broadcasting Tennis Video Clips. In: Proceedings of the 11th International Multimedia Modelling Conference, Melbourne , Australia, 2005. 102～106
51 http://www.ics.uci.edu/~mlearn/MLRepository.html. 2005
52 Lynch R S, Jr, Willett P K. Bayesian classification and feature reduction using uniform Dirichlet priors. IEEE Transactions on Systems, Man and Cybernetics, Part B, 2003, 33(3): 448～464
53 Kantardzic M. 数据挖掘—概念、模型、方法和算法. 闪四清, 陈茵, 程雁, 等译. 北京: 清华大学出版社, 2003
54 Mitchell M T, 等. 机器学习. 北京: 机器工业出版社, 2003
(下转第 12 页)

by conceptual graphs and formal concept analysis. In: Tepfenhart, W, Cyre W, eds. Conceptual Structures: Standards and Practices. Proc. of ICCS '99. LNAI, Springer, Heidelberg, Vol. 1640, 1999. 423~441

16 Obitko M, Snásel V, Smid J: Ontology Design with Formal Concept Analysis. Edited by Vaclav Snasel, Radim Belohlavek. In: Proc. of the CLA 2004 Intl. Workshop on Concept Lattices and their Applications Ostrava, Czech Republic, Sep. 2004. 111~119

17 Chaudhri A F V, Fikes R, Karp P, Rice J. OKBC: A Programmatic Foundation for Knowledge Base Interoperability. In: Proc. of AAAI-98. 1998

18 Haav H M. A Semi-automatic Method to Ontology Design by Using FCA. Edited by Vaclav Snasel, Radim Belohlavek. In: Proc. of the CLA 2004 Intl. Workshop on Concept Lattices and their Applications Ostrava, Czech Republic, Sep. 2004. 13~24

19 Clark H H. Arenas of Language Use, chapter Common Ground and Language Use, Definite Reference and Mutual Knowledge. CSLI, 1992

20 Cimiano P, Stumme G, Hotho A, Tane J. Conceptual Knowledge Processing with Formal Concept Analysis and Ontologies. In: Proc. of the The Second Intl. Conf. on Formal Concept Analysis (ICFCA 04), Springer, 2004. 189~207

21 Cimiano P, Staab S, Tane J. Automatic acquisition of taxonomies from text: FCA meets NLP. In: Proc. of the Intl. Workshop on Adaptive Text Extraction and Mining, 2003

22 Stumme G, M. adche, A. FCA-Merge: bottom-up merging of ontologies. In: Proc. of the Seventeenth Intl. Conf. on Artificial Intelligence (IJCAI '01), Seattle, WA, USA, 2001. 225~230

23 Stumme G, Taouil R, Bastide Y, et al. Lakhal: Computing Iceberg Concept Lattices with Titanic. J. on Knowledge and Data Engineering (KDE), 2002, 42(2): 189~222

24 Ganter B, Stumme G, Creation and merging of ontology top-levels. In: Proc. ECAI02. Submitted, 2002

25 Cole R, Stumme G. CEM: A Conceptual Email Manager. In: 7th Intl. Conf. on Conceptual Graphs, Springer Verlag, ICCS'2000, LNAI 1867, Aug. 2000. 438~453

26 Tane J, Schmitz C, Stumme G, Staab S, Studer R. The Courseware Watchdog: an ontology-based tool for finding and organizing learning material. In Fachtagung "Mobiles Lernen und Forschen". Kassel, Nov. 2003

27 Hotho. Clustern mit Hintergrundwissen: [PhD thesis]. Institute AIFB, University of Karlsruhe, 2004

(上接第 7 页)

55 Naphade M R, et al. A statistical modeling approach to content based video retrieval. In: IEEE Proceedings of 16th International Conference on Pattern Recognition, Quebec, Canada, 2002. 953~956

56 Petkovic M, Jonker W. Content-based video retrieval by integrating spatio-temporal and stochastic recognition of events. In: IEEE Proceedings of IEEE Workshop on Detection and Recognition of Events in Video, Vancouver, Canada, 2001. 75~82

57 Liu T, Kender J R. Semantic mosaic for indexing and compressing instructional videos. In: 2003 International Conference on Image Processing, Barcelona, Spain, 2003. 921~924

58 Dorado A, Calic J, Izquierdo E. A rule-based video annotation system. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(5): 622~633

59 Ho S -Y, et al. Design of Accurate Classifiers With a Compact Fuzzy-Rule Base Using an Evolutionary Scatter Partition of Feature Space. IEEE Transactions on Systems, Man and Cybernetics, Part B, 2004, 34(2): 1031~1043

60 Smeulders A, et al. Content-based image retrieval at the end of the early years. IEEE Trans. On pattern Analysis and Machine Intelligence, 2000, 22(12): 1349~1380

61 M R, et al. TREC 2002 Video Track experiments at MediaTeam Oulu and VTT. In: Proceedings of Text Retrieval Conference TREC 2002 Video Track, Baltimore, MD, USA, 2002. 417~428

62 Luo H, et al. Semantic principal video shot classification via mixture gaussian. In: IEEE Proceedings of 2003 International Conference on Multimedia and Expo, Lausanne, Switzerland, 2003. 189~192

63 Fan J, et al. ClassView: Hierarchical Video Shot Classification, Indexing, and Accessing. IEEE Transactions on Multimedia, 2004, 6(1): 70~86

64 Rasheed Z, Sheikh Y, Shah M. On the use of computable features for film classification. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(1): 52~64

65 Peyrard N, Bouthemy P. Detection of meaningful events in videos based on a supervised classification approach. In: 2003 International Conference on Image Processing, Barcelona, Spain, 2003. 621~624

66 Asano F, Motomura Y, Nakamura S. Fusion of audio and video information for detecting speech events. In: Proceedings of the Sixth International Conference of Information Fusion, Cairns, Australia, 2003. 386~393

67 Miyauchi S, et al. Collaborative multimedia analysis for detecting semantical events from broadcasted sports video. In: 16th International Conference on Pattern Recognition, Quebec, Canada, 2002. 1009~1012

68 Gillespie W J, Nguyen D T. Classification of video shots using activity power flow. In: 2004 IEEE Consumer Communications and Networking Conference, Las Vegas, NV, USA, 2004. 336~340

69 Ekin A, Murat Tekalp A, Mehrotra R. Integrated semantic-syntactic video event modeling for search and retrieval. In: Proceedings, 2002 International Conference on Image Processing, Malaga, Spain, 2002. I-141~I-144

70 Wang J R, et al. Browsing video online using semantic information. In: 7th International Conference on Control, Automation, Robotics and Vision, 2002. 192~197

71 J T, T S, E V. Automatic recognition of emotion in spoken Finnish: preliminary results and applications. In: Proc. Prosodic Interfaces 2003, Nantes, France, 2003. 85~89

72 Hanjalic A, Xu L-Q. Affective video content representation and modeling. IEEE Transactions on Multimedia, 2005, 7(1): 143~154

73 Osadchy M, Keren D. A Rejection-Based Method for Event Detection in Video. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(4): 534~541

74 Ekin A, Tekalp A M, Mehrotra R. Integrated semantic-syntactic video modeling for search and browsing. IEEE Transactions on Multimedia, 2004, 6(6): 839~851

75 Wang F, et al. A Generic Framework for Semantic Sports Video Analysis Using Dynamic Bayesian Networks. In: Proceedings of the 11th International Multimedia Modelling Conference, 2004. MMM 2005, Melbourne, Australia, 2005. 115~122

76 Fan J, et al. Semantic video classification and feature subset selection under context and concept uncertainty. In: Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, 2004, Tuscon, AZ, 2004. 192~201

77 C B, C M. http://www-nlpir.nist.gov/projects/trecvid. 2005

78 JTC1/SC29/WG11 I. I. MPEG-4Overview, Int'l Organization for Standariation. 2002

79 Duch W, Setiono R, Zurada J M. Computational intelligence methods for rule-based data understanding. In: Proceedings of the IEEE, 2004. 85~92

80 Roth B V. Perception and Representation, 2nd ed. Maidenhead: Open Univ. Press, 1995