基于本体的多模式融合语义提取模型

李 敏[1],高 珏[2],吴佳家[1],许华虎[3]
(1. 上海大学 计算机工程与科学学院,上海 200444;
2. 上海大学 计算机中心,上海 200444;
3. 上海上大海润信息系统有限公司,上海 200444)

摘 要:基于语义的视频检索要处理的两项关键技术就是解决视频低层特征和高层语义概念间的语义鸿沟以及有效的语义提取模型。文中通过对视频进行多层次语义分析,采用有效的语义对象分割方法提取视频中的语义对象,以语义对象为中间层,并融合图像、声音、文本的多模式视频特征,从而缩小语义鸿沟。其次,视频语义概念具有多粒度性,由于本体在表示概念及概念间联系时的优越性,文中提出基于本体的语义提取模型,在从图像、声音、文本中提取出的原子概念中,推理出更高层次的复合概念。最终运用该模型提取的视频语义就具有更丰富的语义层次和语义粒度,从而更接近人类思维中的高层语义概念。

关键词:本体;语义提取模型;多模式融合;视频检索
中图分类号:TP301.6 **文献标识码**:A **文章编号**:1673-629X(2013)09-0028-04
doi:10.3969/j.issn.1673-629X.2013.09.007

Multi-mode Fusion Semantic Extraction Model Based on Ontology

LI Min[1],GAO Jue[2],WU Jia-jia[1],XU Hua-hu[3]
(1. College of Computer Engineering and Science,Shanghai University,Shanghai 200444,China;
(2. Computer Center,Shanghai University,Shanghai 200444,China;
3. Shang Da Hai Run Information System Co. Ltd,Shanghai 200444,China)

Abstract:There're two key technologies for semantic based video retrieval has to deal with. One is to bridge the semantic gap between low-level features and high-level semantic concepts. The other is to build a effective semantic model. In this paper,based on the multi-level semantic analysis of video,employ an effective object segmentation method on extracting the semantic object in video,and use semantic object as the middle level and consider the fusion of multi-model,such as image,sound and text,to bridge semantic gap. Meanwhile,videos contain multi-granularity semantic concepts. And since it has superiority to describe concepts and relationships between them,propose a ontology-based semantic model,to reason compound concept of higher level from atomic concepts,which are extracted from image,sound and text. By using this semantic extraction model,can get the video semantics contain richer semantic level and semantic granularity. Thus it can be closer to the semantic concepts in human thoughts.

Key words:ontology;semantic extraction model;multi-model fusion;video retrieval

0 引 言

随着视频数据的海量级增加,迫切需要有效的方法在语义层理解和管理视频数据。基于语义的视频检索[1]关键在于视频语义的定义与理解,以及实现计算机自动提取尽可能与人对视频内容的理解保持一致的视频语义,进而达到视频语义检索的终极目标,即计算机的视频检索能力趋近于人的理解水平。在低层特征向高层语义概念映射时往往产生语义鸿沟,再加之人们在进行视频检索时所使用的查询样本信息并不完善,要实现复杂语义信息查询面临很大的难题。因此,只有解决视频低层特征和高层语义概念间的语义鸿沟、构建有效的语义提取模型,才能实现真正的语义视频检索。为了跨越语义鸿沟,共享视频内容的语义层描述,视频语义检索系统在提取低层特征的基础上,使

收稿日期:2012-11-19 **修回日期**:2013-02-23 **网络出版时间**:2013-04-22
基金项目:上海市科技计划项目(12111101004)
作者简介:李 敏(1988-),女,硕士,研究方向为多媒体技术;高 珏,副教授,研究方向为多媒体、Internet 技术和嵌入式应用;许华虎,教授,博导,研究方向为人机交互、图像处理、多媒体技术等。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20130422.1721.026.html

用本体规范化视频语义概念,建立层次化语义索引成为视频内容分析领域的必然趋势。

1　基于本体的语义概念分析

本体是关于概念体系的,明确的、形式化的,且得到大多数人认同的规范说明。在这个定义的基础上,德国卡尔斯鲁厄大学的 studer 等学者又对本体的特点作了更为明确和直观的解释,即明确的、形式化的、可共享的[2]。本体可以帮助人们获得相关的领域知识,确定在该领域内被共同认可的概念(也即术语)以及这些概念间的相互关系。作为一种有效的知识管理和表示方法,本体已被应用于许多领域,相应的也出现了一些标准本体建模语言,其中比较重要的有 RDF, RDFS, OWL 和针对多媒体应用的 MPEG7 XML Schema。

统计分析法是用视频低层特征的统计特性实现对高层语义概念的检测。当前大多数自动化视频内容分析[3-5]是基于该方法的。由于视频内容的表现形式多样性,仅用低层特征很难有效地概括一些特定的语义概念,采用传统的统计分析法仍无法跨越语义鸿沟。因而应用领域知识将视频高层语义与自动分析获取视频语义的技术集成到统一的框架中是非常必要的。用本体实现对多媒体语义内容在概念级上的建模,以概念本体作为统一术语集实现多媒体语义内容标注、索引以及用户检索概念匹配,构造一个视频查询相关的本体进行视频标注,不仅可以在一定程度上弥补低层特征在视频标注上的不足,还可以充分利用视频的潜在语义特征提高视频内容标注的有效性以及检索的准确性。

2　基于本体和多模式融合的语义检索模型

2.1　视频语义分析

(1)多粒度语义分析。

从人类的认知机理分析,人们思维中的语义是多粒度多层次的,因而视频语义分析也需要统一考虑各层次的语义粒度抽取。根据视频中各语义要素间组合的抽象程度,视频语义层次模型可大致分为特征语义、对象语义、空间关系语义、场景语义、行为语义和情感语义等 6 个层次,用以对不同层次的视频内容进行描述。

视频的语义是多层次多粒度的,而视频检索中,要提取的语义概念大致可划分为两大类:原子概念和复合概念。原子概念是指那些以对象方式存在且不可再分割的概念(如草地、天空、汽车等概念)。复合概念则是那些由多个原子概念组成的、包含多种语义特征的概念(如聚会、人群、法庭等),又或者那些包含抽象语义信息的概念(如学校校长等)。基于低层特征和模式分类的方法尚能较为有效实现对原子概念的检测[6-8],但是对于那些非常具有全局性质的复杂的复合概念就无能为力了。

(2)多模式融合的视频语义分析。

视频具有多模式特征,其每种模式特征都涵盖了丰富的语义信息,而且在不同的应用中既相互独立又相互补充。从生理学上来说,人的各种感觉器官在受到刺激后,便会由中枢神经传递给相应的脑皮层初级区域,使得这些区域中具有相同感受的多种特征检测细胞聚集在一起,从而实现对同一感觉模式的各种刺激的综合反应,形成初步的知觉。而不同模式的感觉信息将由联络区皮层的多模式感知细胞综合为复杂的知觉。人脑的多模式知觉融合恰恰反映了语义理解的层次性,它是简单的两级结构。如果仅从单一模式的特征进行视频语义的分析,那它就只能反映侧面、局部的信息。因此视频分析领域中的一项重要的研究内容便是如何将多模式特征进行有效地融合,从而获取更为全面、准确的高层语义信息。多模式融合的视频语义分析[9]也是跨越语义鸿沟的有效方向。图 1 所示的是人在爆炸现场产生"爆炸"这一语义概念的过程。

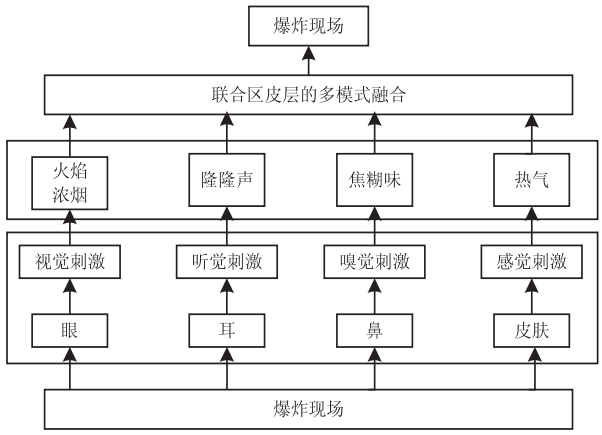

图 1　人脑多模式知觉融合结构

与图 1 所示的人脑多模式知觉融合结构相对应,可这样划分视频中的多模式特征:图像、声音、文本等为第 1 层模式组,每组高层模式中包括的多个同等级别的低层模式特征为第 2 层模式组。视频数据模式划分的结果如图 2 所示。

(3)多层次视频语义分析。

一种有效的层次语义分析方法是基于对象的视频语义分析方法。该方法需要先从视频序列中分割并提取出各种语义对象 。在视频图像中,语义对象包括了前景对象,如运动的汽车、人物、动物等,以及房屋、树木等背景对象。这是视频处理与分析领域中最关键和基础的步骤之一。人们通常是选择性地观看视频,因而在一个视频的所有语义对象中,只有一个或少数能够对人类的感官产生重要影响,这些语义对象就是定

义的关键语义对象[10](目前视频语义对象分割方法[11]非常多且应用丰富,文中采用文献[12]所提出的方法)。如图 3 所示,采用多层分析法,用显著视频语义对象作为中间层,在低层特征与高层语义之间构建了一座桥梁,从而跨越了语义鸿沟,得到人类思维中的概念。

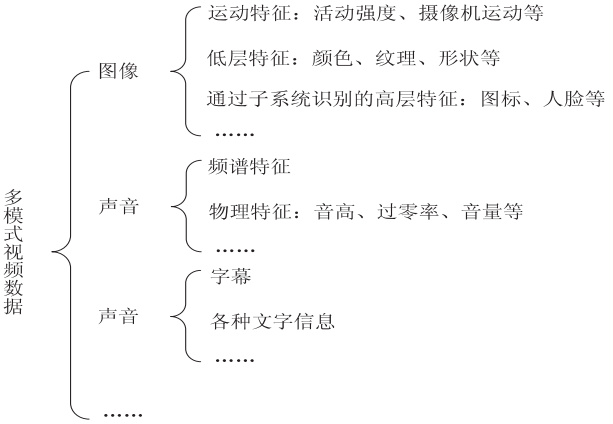

图 2　视频的多模式划分

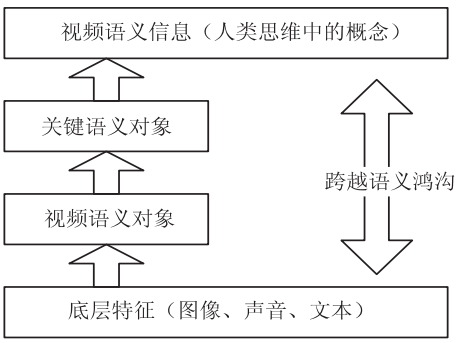

图 3　跨越语义鸿沟的框架

2.2　基于本体和多模式融合的语义提取模型

(1)基于本体的语义提取模型。

本体所描述的概念是较为规范且得到一致认可的,因而采用本体来表示语义是合理有效的。文中利用本体中的概念对视频进行标注,将会更适用于语义视频检索的研究,也更加规范和通用。运用本体理论,可以从全新的角度而非传统的基于低层特征的模式分类方法考虑视频标注和检索问题。通过对领域知识架构及规则的充分利用,可以从语义推理的角度结合基于低层特征的模式识别和基于领域知识的概念推理,实现更加准确的视频语义特征提取。

本体和分类学是两种主要的语义模型的建立方法。分类学方法往往只定义了概念之间的层次关系,因而它在描述概念之间错综复杂的关系时显得十分的简单[13]。相反,本体能从不同层次的形式化模式上明确定义概念间的关系,因而成为构建语义模型的重要方法,并普遍应用于各种复杂的应用中。

文中采用的语义模型可记为 O,其表示形式为:$O=<C,A,R,I,M>$。其中,C、A、R、I、M 分别表示概念集、属性集、关系、实例集和实例与概念之间的映射关系集合。概念集是指特定领域中概念、术语的集合;属性集为概念自身的特征;实例集表示各概念间的交互作用;映射关系集合则将每个实例对应到其所属的概念下。例如,对 I 中的一个实例 i,若有 $i \in M(c)$,则说明 i 是概念 c 的实例。

图 4 为一个本体语义模型结构图。图中的圆形表示概念,而箭头则表示概念间的关系。其中,实例属于相应的概念,而各个概念间通过关系彼此关联。用户要检索的信息便以实例的形式与概念建立起联系。

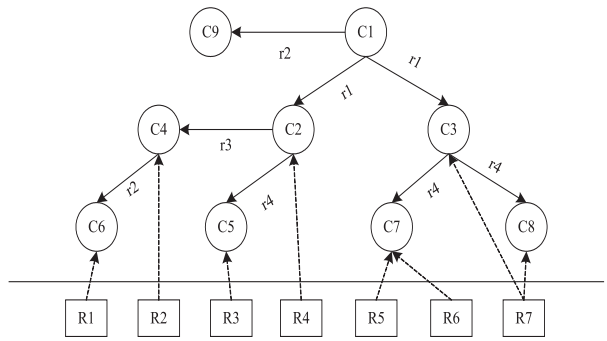

图 4　本体语义模型结构

将语义检索模型记为 N,它是一个非确定型有穷自动机,则有 $N=<Q,\Sigma,\delta,q_0,F>$。$Q$、$\Sigma$、$\delta$、$q_0$、$F$ 分别表示概念集、关系集、概念转移函数、语义检索的起始概念以及接受状态集,其中 Q、Σ 为有穷集。Q 的一个幂集 $\rho(Q)$,$Q\times\Sigma\rightarrow\rho(Q)$ 为 δ 中的概念转移函数。q 为 Q 中的一个概念,r 为 Σ 中的一个关系,经过转移函数 δ 的作用,总有 $\delta(q,r)\subseteq F$。

在表 1 中,定义了图 4 所示的语义模型中的部分概念转移函数。以 $C3$ 作为起始概念,当读入关系 $r3$ 时,则转变到概念集 $\{C7,C8\}$;读入任何其他关系都将转到空集 $\varnothing$。

表 1　语义模型的部分概念转移函数

概念	关系 $r1$	关系 $r2$	关系 $r3$	关系 $r4$
$C1$	$\{C3,C1\}$	$\{C9\}$	$\varnothing$	$\varnothing$
$C2$	$\varnothing$	$\varnothing$	$\{C4\}$	$\{C5\}$
$C3$	$\varnothing$	$\varnothing$	$\{C7,C8\}$	$\varnothing$
$C4$	$\varnothing$	$\{C6\}$	$\varnothing$	$\varnothing$

(2)语义提取总体框架。

结合 2.1 节中的分析,文中提出了一种基于本体的、以语义对象作为中间层、考虑视频多模式信息的视频语义提取模型。其核心是构造一个概念检测本体,在对象语义级别上同时构造对象、声音、文本三种模式的原子概念,然后利用概念检测本体,检测复合概念与原子概念之间的依赖关系,从而推断出视频的复合概念,也就实现了从视频对象层语义到视频场景语义或行为语义的推理。该语义提取框架如图 5 所示。

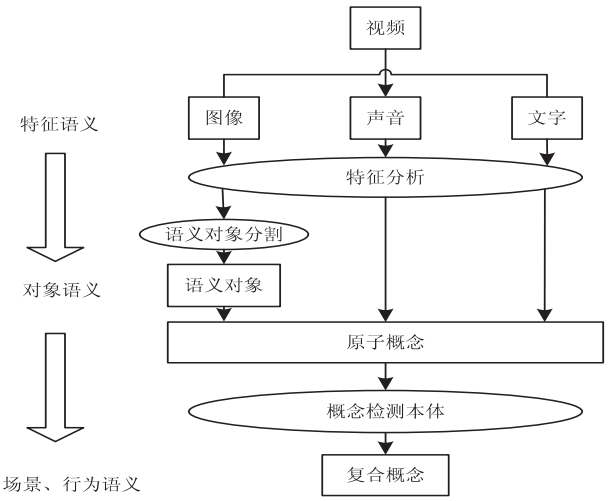

图 5　语义提取框架

3　结束语

文中提出了一种新的基于本体的复合概念检测模型。不仅利用了本体表示概念的优越性,同时也考虑了视频语义的多层级多粒度性,以及视频的多模式特征,扩展和加深了语义模型对视频语义的分析和理解,使得自动、全面的视频语义分析得以更好的实现。

参考文献:

[1] 魏　纬,游　静,刘凤玉,等. 语义视频检索综述[J]. 计算机科学,2006,33(2):1-7.

[2] 李　景. 本体理论在文献检索系统中的应用研究[M]. 北京:北京图书出版社,2005.

[3] 王　忠. 基于内容的视频检索关键技术的研究与实现[D]. 西安:西安电子科技大学,2009.

[4] 史迎春,周献中,方鹏飞. 综合利用形状和颜色特征的台标识别[J]. 模式识别与人工智能,2005,18(2):216-222.

[5] 桂丹萍,陈佳祥,何红生. 视频检索在汉字识别中的应用研究[J]. 计算机技术与发展,2010,20(10):207-210.

[6] Natsev A,Naphade M R,Smith J R. Semantic Representation Search and Mining of Multi-media Content[C]//Proceeding of ACM International Conference on Knowledge Discovery and Data Mining. Seattle,WA,USA:[s. n.],2004:641-646.

[7] Wu Y,Chang E Y. Optimal multi-model fusion for multimedia data analysis[C]//Proceeding of ACM International Conference on Multimedia. New York,USA:[s. n.],2004:572-579.

[8] Hauptmann A,Chen M Y,Christel M,et al. Confounded Expectations:Informedia at TRECVID 2004[C]//Proc. of NIST TRECVID Workshop. Gaithersburg,USA:[s. n.],2004.

[9] 魏　纬,邹书蓉,刘凤玉. 多层视频语义概念分析与理解[J]. 计算机辅助设计与图形学学报,2008,20(1):85-92.

[10] 张　良. 一种视频关键语义对象的检测方法[J]. 北京信息科技大学学报,2010,25(2):75-78.

[11] 刘　志,张兆杨. 语义对象分割技术综述[J]. 上海大学学报(自然科学版),2007,13(4):477-484.

[12] Li Min,Gao Jue,Wu Jiajia,et al. Semantic Object Segmentation in Video Retrieval[C]//Proc. of the 2nd International Conference on Electronics,Communications and Control. [s. l.]:IEEE,2012.

[13] Ganong W F. Review of Medical Physiology[M]. New York:McGraw-Hill Publishing Company,1999.

(上接第 27 页)

[4] Li J,Blumenfeld D E,Huang N,et al. Throughput analysis of production systems:recent advances and future topics[J]. International Journal of Production Research,2009,47(14):3823-3851.

[5] Jacobs D A,Meerkov S M. Mathematical theory of improvability for production systems[J]. Mathematical Problems in Engineering,1995,1(2):95-137.

[6] Chiang S Y,Kuo C T,Meerkov S M. c-Bottleneck in serial production lines[J]. Mathematical Problems in Engineering,2001,7:543-578.

[7] Li J. Performance analysis of production systems with rework loops[J]. IIE Transactions,2004,36(8):755-765.

[8] Li J. Overlapping decomposition:a system-theoretic method for modeling and analysis of complex production systems[J]. IEEE Transactions on Automation Sciences and Engineering,2005,2(1):40-53.

[9] Liu Y,Li J,Chiang S Y. Re-entrant lines with unreliable asynchronous machines and finite buffers:performance approximation and bottleneck identification[J]. International Journal of Production Research,2011,50(4):977-990.

[10] 幸　研,易　红,汤文成. 制造系统工作流设计的校验和性能分析方法[J]. 西安交通大学学报,2002,36(3):278-281.

[11] 侯　扬,范秀敏,严隽琪,等. 基于仿真的制造系统对象建模及其应用[J]. 计算机集成制造系统,2001,7(5):42-46.

[12] 卫军胡,韩九强,孙国基. 离散事件系统仿真技术在制造系统调度中的应用[J]. 系统仿真学报,2000,12(1):27-30.

[13] 吴文涛,朱华炳. 基于 UML 的桥壳生产系统仿真建模研究[J]. 组合机床与自动化加工技术,2010(2):95-97.

[14] 刘　颖,王　伟,刘全利. 基于 SystemC 的罩式炉退火车间离散事件仿真模型[J]. 大连理工大学学报,2010,50(1):145-151.

[15] 范金松,严洪森,周久海,等. 基于遗传算法的某航空发动机装配车间优化调度[J]. 计算机技术与发展,2012,22(9):205-209.

[16] Kumar P R. Re-entrant lines[J]. Queueing Systems,1993,13(1-3):87-110.

基于本体的多模式融合语义提取模型

万方数据 WANFANG DATA 文献链接

作者： 李敏， 高珏， 吴佳家， 许华虎， LI Min， GAO Jue， WU Jia-jia， XU Hua-hu

作者单位： 李敏,吴佳家,LI Min,WU Jia-jia(上海大学 计算机工程与科学学院,上海,200444)， 高珏,GAO Jue(上海大学 计算机中心,上海,200444)， 许华虎,XU Hua-hu(上海上大海润信息系统有限公司,上海,200444)

刊名： 计算机技术与发展

ISTIC

英文刊名： Computer Technology and Development

年，卷(期)： 2013(9)

参考文献(13条)

1. 魏纬. 游静. 刘凤玉 语义视频检索综述[期刊论文]-计算机科学 2006(02)

2. 李景 本体理论在文献检索系统中的应用研究 2005

3. 王忠 基于内容的视频检索关键技术的研究与实现 2009

4. 史迎春. 周献中. 方鹏飞 综合利用形状和颜色特征的台标识别[期刊论文]-模式识别与人工智能 2005(02)

5. 桂丹萍. 陈佳祥. 何红生 视频检索在汉字识别中的应用研究[期刊论文]-计算机技术与发展 2010(10)

6. Natsev A. Naphade M R. Smith J R Semantic Representation Search and Mining of Multi-media Content 2004

7. Wu Y. Chang E Y Optimal multi-model fusion for multimedia data analysis 2004

8. Hauptmann A. Chen M Y. Christel M Confounded Ex-pectations:Informedia at TRECVID 2004 2004

9. 魏纬. 邹书蓉. 刘凤玉 多层视频语义概念分析与理解[期刊论文]-计算机辅助设计与图形学学报 2008(01)

10. 张良 一种视频关键语义对象的检测方法[期刊论文]-北京信息科技大学学报 2010(02)

11. 刘志. 张兆杨 语义对象分割技术综述[期刊论文]-上海大学学报（自然科学版） 2007(04)

12. Li Min. Gao Jue. Wu Jiajia Semantic Object Segmenta-tion in Video Retrieval 2012

13. Ganong W F Review of Medical Physiology 1999

本文链接：http://d.g.wanfangdata.com.cn/Periodical_wjfz201309007.aspx