# 基于本体的视频语义内容分析

白 亮 刘海涛 老松杨 卜 江
(国防科学技术大学信息系统与管理学院 长沙 410073)

**摘 要** 随着视频数据的大量涌现,迫切需要有效的方法在语义层理解和管理视频数据。新的多媒体标准,如MPEG-4、MPEG-7等,对操纵和传输视频对象及元数据提供了基本的功能框架。但重要的是,视频数据的语义层内容大部分超出了标准涉及的范围。提出了一个基于本体的视频语义内容分析框架,采用领域本体定义目标领域中的高层语义概念及语义概念在上下文间的关系;为增强视频语义分析能力,将低层特征(如视觉和听觉)和视频内容分析算法集成进本体中;采用OWL(Web Ontology Language)作为本体建模语言;根据不同的感知内容和低层特征,定义描述逻辑(Description Logic,简称DL)描述不同的视频特征和处理算法如何应用于应用视频分析;采用时域描述逻辑(Temporal Description Logic,简称TDL)来描述语义事件,并且提出一个推理算法进行事件探测。提出的框架在足球视频领域进行了实验验证,得到了令人满意的实验结果。

**关键词** 本体,语义内容分析,语义事件
**中图法分类号** TP37 **文献标识码** A

## Video Semantic Content Analysis Based on Ontology

BAI Liang LIU Hai-tao LAO Song-yang Bu Jiang
(School of Information System & Management, National University of Defense Technology, Changsha 410073, China)

**Abstract** The rapid increase in the available amount of video data is creating a growing demand for efficient methods for understanding and managing it at the semantic level. New multimedia standards, such as MPEG-4 and MPEG-7, provide the basic functionalities in order to manipulate and transmit objects and metadata. But importantly, most of the content of video data at a semantic level is out of the scope of the standards. A video semantic content analysis framework based on ontology was presented. Domain ontology was used to define high level semantic concepts and their relations in the context of the examined domain. And low-level features(e. g. visual and aural) and video content analysis algorithms were integrated into the ontology to enrich video semantic analysis. OWL was used for the ontology description. Rules in Description Logic were defined to describe how features and algorithms for video analysis should be applied according to different perception content and low-level features. Temporal Description Logic was used to describe the semantic events, and a reasoning algorithm was proposed for events detection. The proposed framework was demonstrated in a soccer video domain and shows promising results.

**Keywords** Ontology, Semantic content analysis, Semantic event

## 1 引言

随着高速宽带网络、数字视频和硬件技术的迅速发展,视频已成为WWW、数字电视、数字图书馆和视频点播等多媒体应用领域的重要内容源。视频资源的快速增长,迫切需要智能的方法在语义层对视频数据进行理解、存储、索引和检索[1]。

一方面,新的多媒体标准(如MPEG-4、MPEG-7等)为操纵和传输对象及元数据提供了基本的方法,但是在语义层面上,大部分视频内容都超出了标准的适用范围。另一方面,特征提取、镜头探测和对象识别是视频内容分析的重要步骤[2,3],过去20年出现了一些重要的研究成果并出现了几个成功的原型系统[4-6]。然而,在视频语义内容的表示方面缺乏精确的模型和规范化形式,同时视频处理算法的高度复杂性也使得开发完全自动化的视频语义分析管理系统成为目前具有挑战性的任务。

视频内容分析领域的主要挑战是如何建立视频高层语义内容和低层时空特征间的关联,即经典的语义鸿沟问题。在许多应用系统中,映射规则必须固化进程序代码,这导致现有方法和系统太"僵硬",缺乏灵活性,不能满足语义层上的视频应用。因此,有必要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。

到稿日期:2008-08-04 返修日期:2009-10-05 本文受863项目(2006AA01Z316),国家自然科学基金(60572137)教育部博士点基金项目资助。
白 亮(1978—),男,博士研究生,主要研究领域为数字视频分析与处理技术,E-mail:xabpz@163.com;刘海涛(1978—),男,博士研究生,主要研究领域为数字视频分析与处理技术;老松杨(1968—),男,教授,博士生导师,主要研究领域为数字视频处理和检索技术;卜 江(1983—),男,博士研究生,主要研究领域为数字视频分析与处理技术。

本体是领域知识确定的、形式化的规范描述，是一种有效的语义建模和知识表示工具，由概念、概念属性、概念间的关系组成，通常使用语言术语描述。本体作为知识管理和表示的方法已经应用于许多领域，同时出现了一些标准本体建模语言，其中比较重要的有 Resource Description Framework (RDF)[7]，Resource Description Framework Schema(RDFS)，Web Ontology Language(OWL)[8] 和针对多媒体应用的 MPEG-7 XML Schema。

近年来出现了许多自动语义内容分析系统，文献[9]通过 HMM 模型建模 MPEG 运动向量、球场形状和运动员位置来探测足球中的精彩部分；A. Ekin 等人探测慢镜头的出现来推断体育视频中的精彩部分[10]；文献[11]中用球的运动轨迹探测主要动作并计算每个队的控球率；文献[12]中设计了一个框架，利用内部视音频特征和扩展信息来探测体育视频中的事件；Sadlier 等人[18]探测视频低层和中层特征并用支持向量机来探测足球视频中的精彩片断。这些系统都采用基于低层特征的语义内容分析方法，没有利用任何领域知识的形式化描述。

常用的语言本体的形式化描述是基于语言术语的。文献[13]提出集成多媒体词典和文档交叉融合概率的领域语言本体；文献[14]中将概念表示为关键词并映射到一个对象本体、一个镜头本体和一个语义本体来作为视频分割结果的表示。然而，虽然语言术语适合于区分给定领域中的事件和对象类别，但在描述低层特征、视频内容分析以及它们之间的关系上存在着困难。

文献[15]提出使用多媒体本体扩展语言本体的方法，以支持视频理解；文献[16]中采用人工构建多媒体本体提取视频中可利用的文本信息和视觉信息并人工组织本体中的概念、属性和联系；Marco Bertini 等人在文献[17]中提出了使用增强本体进行视频标注和检索的算法和技术，提出一个无监督的聚类算法来创建可视化的增强本体，定义代表精彩事件特殊模式的视觉属性并将它们作为视觉概念加入到本体中；文献[19]提出了一个基于本体基础架构的视频语义内容分析和标注方法，并使用 RDF/RDFS 建模领域知识本体；文献[20]分别基于 MPEG-7 视觉描述符和 MPEG-7 多媒体描述框架构建了视觉描述本体和多媒体结构本体，并与领域本体集成，以支持视频内容标注；文献[21]引入一个对象本体，结合相关反馈机制以建立低层特征到高层特征的映射，并且允许定义多媒体信息段之间的关系。

上述研究工作的共同点是利用本体在概念级建模多媒体语义内容，使用概念本体作为多媒体语义内容标注、索引和用户检索概念匹配的统一术语集，以提高多媒体内容标注的有效性和检索的准确性。但是，如何扩展语言本体的表达能力，使之适应建模视频内容包含的多通道语义内容和复杂关系；如何将本体的语义表达能力和推理技术应用到视频语义内容自动探测，都是目前面临的主要挑战。本文提出了一个基于本体的视频语义内容分析框架，构建视频分析本体来形式化描述视频语义内容的探测步骤，将目标领域上下文中的语义概念定义到一个领域本体中；定义描述逻辑规则来描述如何根据不同的感知内容和低层特征选择应用于视频分析的特征和算法；采用时域描述逻辑描述语义事件并提出一个推理算法进行事件探测。使用 OWL 语言建模本文提出的本体框架，通过应用本体建模的领域知识，分析目标领域中的视频语义内容，在语义层面支持内容标注和事件探测。本文第 1 节提出基于本体的视频内容分析总体框架；第 2 节中给出本体的详细定义；第 3 节基于所定义的本体构建应用于视频处理和事件探测的描述逻辑规则；第 4 节解释如何将本文提出的框架应用于一个特定领域——足球视频领域；第 5 节是实验结果与分析；最后总结全文，并提出未来的研究方向。

## 2 基于本体的视频语义内容分析框架

本文提出的视频语义内容分析框架如图 1 所示。视频分析本体是视频分析知识的抽象和形式化描述。通常，视频内容探测需要定义合适的低层特征，使用合适的探测算法，考虑探测过程中必要的约束(例如算法参数、特征的阈值等)。所有这些要素通过视频分析本体统一描述建模，使得视频分析知识具有规范统一的描述，可以共享和重用，支持自动化的视频语义内容分析过程。构建领域本体描述目标领域中的语义概念，同时集成视频语义内容的性质属性、低层特征和视频语义内容及其低层特征决定的视频处理算法，增强领域本体对视频语义内容分析和描述的表达能力。用 OWL 语言描述视频分析本体和领域本体。定义 DL 规则来描述如何根据不同的感知内容和低层特征选择应用于视频分析的特征和算法，以探测特殊语义对象和本体中定义的相应高层语义概念的序列。TDL 能够建模时域关系并定义领域中的重要语义事件，基于 DL 和 TDL 的推理能够支持对象、序列和事件的自动探测。

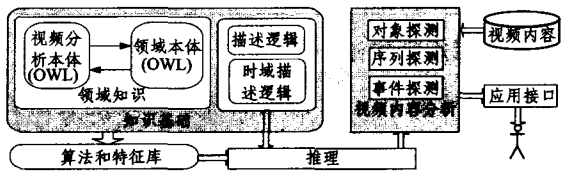


图 1 基于本体的视频语义内容分析框架

基于此框架，视频语义内容分析依赖于系统的知识基础。如果知识基础增加了相应的领域本体，此框架就能很容易应用在不同的领域。此外，基于本体的方法和应用 OWL 语言也能保证提出的框架可以很好地集成到语义 Web 服务和应用中，用于发现和利用视频数据中抽取的信息和知识。

## 3 建立视频语义内容分析本体

为了实现上面介绍的基于知识的自动视频语义内容分析，需要抽象出视频分析知识并构建一个视频分析本体。在视频内容分析领域已有许多应用成熟的特征和算法。通常，视频内容(如对象)探测需要根据已有的算法和特征，考虑使用区分能力强的特征、选择合适的探测算法，应用于内容分析过程。因此，视频内容分析中的所有元素，包括内容、特征、算法和必要的约束，必须在视频分析本体中清楚地描述。本文提出的视频分析本体主要建模如下的概念(OWL 类)和相应的属性。

事件类(Class Event)：超类“Event”的子类和实例。每个事件实例是特定对象实例、序列实例及其时序关系的组合。

序列类(Class Sequence)：超类“Sequence”的子类和实例。序列类可以分为视觉序列和听觉序列子类。通过镜头层的分析处理能够对所有视觉序列进行分类。例如体育视频中的远

景镜头和近景镜头。每个序列实例通过 hasFeature 属性与适当的特征实例相关联，通过 useAlgorithm 属性与适当的算法实例相关联。

对象类(Class Object)：超类“Object”的子类和实例。对象类包含视觉对象子类和听觉对象子类。通过视频帧层的分析处理能够探测视觉对象。每个对象实例通过 hasFeature 属性与适当的特征实例相关联、通过 useAlgorithm 属性与适当的算法实例相关联。

特征类(Class Feature)：与每个序列和对象相关联的视频低层特征的超类，包括音频低层特征。

特征参数类(Class FeatureParameter)：表示每个相应特征的定量描述，可以根据已定义的特征来划分子类。

参数值域类(Class pRange)：包括 Minimum 和 Maximum 两个子类。允许对不同的特征参数定义约束值。

算法类(Class Algorithm)：视频分析过程中可利用的算法的超类。它通过 useFeatureParameter 属性与特征参数类的实例相联系。

本文采用 OWL 来描述上述定义类。视觉对象类的一个描述示例如表 1 所列。

表 1　用 OWL 描述的视觉对象类示例

```
〈owl:Class rdf:ID="VisualObject"〉
    〈rdfs:subClassOf rdf:resource="#Object" /〉
    〈owl:onProperty〉
        〈owl:FunctionalProperty rdf:about="#hasFeature"〉
        〈owl:FunctionalProperty rdf:about="#useAlgorithm"〉
    〈/owl:onProperty〉
〈/owl:Class〉
```

## 4　描述逻辑规则构建

如第 2 节所述，在视频内容分析研究领域已经提出了许多特征和算法。序列和对象探测算法的选择依赖于可利用的特征，而特征又直接依赖于序列和对象所涉及的领域，因此这种关系应该基于视频分析知识和领域知识来考虑。并且这对自动、准确探测序列和对象也是有价值的。本文定义了一个适当的规则集来描述如何选择应用于视频分析的特征和算法，规则集中的规则用 DL 表示。

序列和对象探测的规则为：用来定义序列(或对象)和特征之间映射的规则、用来定义序列(或对象)和算法之间映射的规则及确定算法的输入特征参数的规则。规则表示如下：

• 一个序列‘S’具有特征 $F_1, F_2, \cdots, F_n$：$\exists hasFeature(S, F_1, F_2, \cdots, F_n)$

• 采用算法 $A_1, A_2, \cdots, A_n$ 探测一个序列‘S’：$\exists useAlgorithm(S, A_1, A_2, \cdots, A_n)$

• 一个对象‘O’具有特征 $F_1, F_2, \cdots, F_n$：$\exists hasFeature(O, F_1, F_2, \cdots, F_n)$

• 采用算法 $A_1, A_2, \cdots, A_n$ 探测一个对象‘O’：$\exists useAlgorithm(O, A_1, A_2, \cdots, A_n)$

• 算法‘A’使用特征参数 $FP_1, FP_2, \cdots, FP_n$：$\exists useFeatureParameter(A, FP_1, FP_2, \cdots, FP_n)$

• 如果 $S \cap (\exists hasFeature.F \cap \exists hasAlgorithm.A)$，则 $\exists useFeatureParameter(A, FP)$($FP$ 为 $F.$ 的参数值)

• 如果 $O \cap (\exists hasFeature.F \cap \exists hasAlgorithm.A)$，则 $\exists useFeatureParameter(A, FP)$ ($FP$ 为 $F.$ 的参数值)

下一节将构建一个领域本体，该本体提供了领域词汇和背景知识。在视频内容分析背景中，将领域本体映射到重要对象、对象的定量及定性属性和它们之间的相互关系。

视频中的事件是非常重要的语义内容。事件通常由特定的对象、序列及其时域关系组成。通用的领域本体适合于用语言术语描述事件，但它无法有效地描述事件的时域模式。基本的 DL 缺乏能够表示时间语义的构造符，因此本文采用 TDL 来描述语义事件的时域模式。TDL 是对 DL 的时间扩展，能够描述线性的、无约束的、离散的时域结构。本文使用 TL-F 作为基本的逻辑描述语言。此语言由能够表达时域间隔网络的时间逻辑 TL 和非时域的特征描述逻辑 F 组成[22]。

TL-F 中基本的时域间隔关系包括：*before*(*b*)，*meets*(*m*)，*during*(*d*)，*overlaps*(*o*)，*starts*(*s*)，*finishes*(*f*)，*equal*(*e*)(如图 2 所示，$i$ 和 $j$ 为时域间隔)。

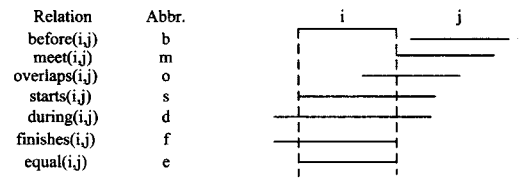


图 2　两个时域间隔间的关系

足球视频中的对象和序列能够用一个视频分析本体来探测，事件可以依赖视频中对象和序列的出现及它们之间的时域关系来描述。下一节将具体介绍事件探测所用的事件描述和推理算法。

## 5　足球领域本体

为了验证所提出的框架，本文实现了一个足球领域的应用。重要语义序列和对象(如特写镜头、球员和裁判等)的探测对理解和提取视频语义内容、建模和探测视频中的事件是非常重要的。与序列和对象关联的特征根据视频分析背景中的低层特征进行定义。序列和对象的类别以及特征的选择根据领域知识决定。通过分析观察和对足球领域知识的总结学习，本节描述了足球领域本体的构建和定义。

### 5.1　对象

通过观察，足球视频中仅包含有限数量的对象类型，视觉对象包括足球、球员、裁判(助理裁判)、教练、球门、边线、角弧和屏幕字幕。根据语义内容分析的具体需求，本文定义 3 个对象作为独立的视觉对象类实例：字幕、球门和裁判(如图 3 所示)。并对对象的探测镜头和探测位置进行了如下限制，以减小探测难度。

(a) 字幕　(b) 球门　(c) 裁判

图 3　足球视频中的对象

• 字幕：探测下屏字幕，但不要求进行字幕定位与识别。

• 球门：在场中远焦镜头和慢镜头中探测球门。

• 裁判：在特写镜头中探测裁判。

比赛中的一些声音对视频语义分析具有重要作用。通常，在足球比赛中有两种重要的音频：哨声和欢呼声。因此本文定义两个独立的听觉对象类实例：哨声和欢呼声。

### 5.2 序列

足球视频中通常包含3种明显的视觉序列类：远景视角序列、中景视角序列和近景视角序列（如图4中(a)、(b)，(c)所示）。


图4 足球视频中的序列

足球视频中远景序列和中景序列具有相似的视觉特征，并经常出现在同一个镜头变焦动作中，因此将它们定义为一种序列类型，称为普通视角序列。当一些精彩事件发生，相机通常会捕捉该区域的一些有趣事件，即所谓的场外事件。发生重要的语义事件后，通常会立即伴随着慢镜头的回放。因此，定义独立的视觉序列类为普通视角序列（Normal View，简称NV）、近景视角序列（Tight View，简称TV）、场外视角序列（Out-of-field，简称OOF）和慢镜头序列（Slow-motion-replay，简称SMR）。基于球场的不同区域，可以进一步将普通视角划分为8个子类（the "pitch"，如图5所示）：左球门区域、右球门区域、上中场、下中场、左上角、左下角、右上角和右下角。

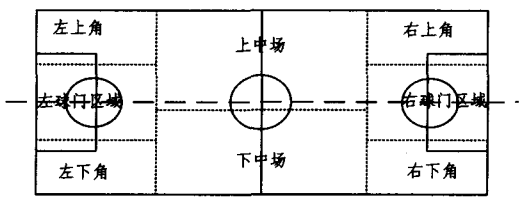


图5 足球场的8个区域

### 5.3 特征和算法

通过研究足球领域中对象和序列的特点，发现能够通过相似的颜色或形状来区分足球视频中的视觉对象和序列，因此在足球领域本体中应用颜色特征和形状特征，采用MPEG-7颜色和形状特征描述符[27]。

主颜色（Dominant Color）和颜色分布（Color Layout）是颜色特征中的两个独立特征。在使用少数的颜色就能够表现感兴趣区域的颜色特征的情况下，主颜色能够有效地描述局部特征；颜色分布对于描述颜色的空间分布是有效的。颜色特征对于区分不同的视觉序列和探测具有相似颜色特征的视觉对象是有效的，如"裁判"等。区域形状特征（Region Shape Feature）能够描述复杂的形状，并且对于对象的边界局部失真具有鲁棒性。形状特征对于探测足球视频中具有固定形状的对象是有效的，如"字幕"。

在先前的工作中[23]，采用HMM来区分不同的视觉序列；用Sobel边界探测和Hough变换来探测"球门"对象。基于颜色特征的图像聚类算法已被证明在足球视频内容分析领域是有效的。并通过计算视频流中基于像素的平均平方亮度差值，采用多门限过零率方法探测慢镜头[24]。对于听觉对象的探测，将频域能量应用于SVM模型来探测"欢呼"[25]，并根据频率峰值与设定阈值范围的比较来探测"哨声"[26]。

### 5.4 事件描述和探测

采用视频内容分析本体，探测上文定义的序列和对象。基于序列和对象探测结果，可以探测足球视频中的事件。我们根据一些探测到的序列和对象观察足球视频中的时域模式。例如，如果一次进攻导致进球事件，观众的欢呼声会立即出现，同时相应视觉序列和对象的时域模式表现为从"球门区域"到"球员特写"、"场外区域"、"慢镜头回放"和其他球员的"特写"，最后返回"普通视角"，然后出现"字幕"。本质上，这些时域模式是足球视频领域中常用的编辑和拍摄手法的体现，可以看作是领域中基本的事实，即领域知识。它描述了足球视频中语义事件的特征，并能够用来形式化地建模语义事件和自动探测事件。如第3节的论述，本文采用TDL来进行事件形式化描述，并以进球和犯规事件为例说明本文提出的方法。采用TDL中的必要的语法和符号解释如下：

$x, y$ 表示时域间隔；

◇是一个时域存在量词，用来引入时域间隔，如◇$(x, y)$；

@称为约束，出现在时域间隔的左边。一个约束变量用来"绑定"一个概念，表示被"绑定"的概念出现在最近的时域间隔范围内。

·进球

足球视频中，一个进球事件通常包含球对象、哨声对象、欢呼对象、字幕对象和球门区域（GA）、TV序列、OOF序列、SMR序列。用TDL描述进球事件如下所示：

$$Scoredgoal = \Diamond(d_{goal}, d_{whistle}, d_{cheers}, d_{caption}, d_{GA}, d_{TV}, d_{OOF}, d_{SMR})(d_{goal}\, fd_{GA})(d_{whistle}\, dd_{GA})(d_{GA}\, od_{cheers})(d_{caption}\, ed_{TV})(d_{cheers}\, ed_{TV})(d_{GA}\, md_{TV})(d_{TV} md_{OOF})(d_{OOF} md_{MSR}).$$

$$(goal@d_{goal} \cap whistle@d_{whistle} \cap cheers@d_{cheers} \cap caption@d_{caption} \cap GA@d_{GA} \cap TV@d_{TV} \cap OOF@d_{OOF} \cap SMR@d_{SMR})$$

$d_{goal}, d_{whistle}, d_{cheers}, d_{caption}, d_{GA}, d_{TV}, d_{OOF}, d_{SMR}$ 表示相应对象和序列的时域间隔。

·犯规

足球视频中，犯规事件通常发生在一个带有哨声对象的NV序列中，接着是一个带有裁判对象的TV序列，最后是一个MSR序列。用TDL描述犯规事件如下所示：

$$Foul = \Diamond(d_{whistle}, d_{referee}, d_{NV}, d_{TV}, d_{SMR})(d_{whistle}\, dd_{NV})(d_{referee}\, dd_{TV})(d_{NV}\, md_{TV})(d_{TV}\, md_{MSR}).$$

$$(whistle@d_{whistle} \cap referee@d_{referee} NV@d_{NV} \cap TV@d_{TV} \cap SMR@d_{SMR})$$

$d_{whistle}, d_{referee}, d_{NV}, d_{TV}, d_{SMR}$ 表示相应对象和序列的时域间隔。

如果犯规导致一个红牌或黄牌，则会出现一个字幕对象，红牌或黄牌事件的描述如下所示：

$$Foul = \Diamond(d_{whistle}, d_{referee}, d_{caption}, d_{NV}, d_{TV}, d_{SMR})(d_{whistle}\, f\, d_{NV})(d_{referee}\, d\, d_{TV})(d_{caption}\, d\, d_{TV})(d_{NV}\, m\, d_{TV})(d_{TV}\, m\, d_{MSR}).$$

$$(whistle@d_{whistle} \cap referee@d_{referee} \cap caption@d_{caption} NV@d_{NV} \cap TV@d_{TV} \cap SMR@d_{SMR})$$

$d_{caption}$ 为字幕对象的时域间隔。

基于TDL的事件描述，本文提出了一种用于事件探测的推理算法。在探测出足球视频中的序列和对象后，用TDL正式描述每个序列和对象如下：

$\Diamond x().C@x$

$C$ 为单独的序列或对象；$x$ 为 $C$ 的时域间隔。()表示 $C$ 与其

自身没有任何时域关系。

推理算法描述如下：

假设 $\{S_0,S_1,\cdots,S_{n-1},S_n\}$是从一个足球视频中探测得到的序列集，$\{S_0,S_1,\cdots,S_{n-1},S_n\}$中的每个元素 $S_i$ 都能用下式表示：

$S_i=\diamondsuit x_i().\ S_i@x_i$

$\{S_0,S_1,\cdots,S_{n-1},S_n\}$的定义包含一个潜在的时域约束：$x_i\ m\ x_{i+1}, i=0,1,\cdots,n-1$ 表示$\{S_0,S_1,\cdots,S_{n-1},S_n\}$中两个连续的序列在视频的时间轴上是连续的。

$\{O_0,O_1,\cdots,O_{m-1},O_m\}$是从一个足球视频中探测得到的对象集。$\{O_0,O_1,\cdots,O_{m-1},O_m\}$中的每个元素 $O_i$ 都能用下式表示：

$O_i=\diamondsuit y_i().\ O_i@y_i$

以进球事件为例，事件探测推理算法可描述如下：

Step 1 选择$\{S_0,S_1,\cdots,S_{n-1},S_n\}$中包含连续序列 GA->TV->OOF->MSR 的子集，每个子集都是一个候选的进球事件 $E_{Gk}$：

$E_{Gk}=\{GA_k,TV_{k+1},OOF_{k+2},MSR_{k+3}\}$

其中 $k$ 是$\{S_0,S_1,\cdots,S_{n-1},S_n\}$中当前候选事件中的当前 NV 的下标。

Step 2 对于每个候选事件 $E_{Gk}$，在$\{O_0,O_1,\cdots,O_{m-1},O_m\}$中搜索对象 $O_{goal}$，$O_{whistle}$，$O_{cheers}$，$O_{caption}$，它们对应的时域间隔分别为 $y_{goal}$，$y_{whistle}$，$y_{cheers}$，$y_{caption}$，满足相应的时域约束 $y_{goal}\ f\ GA_k$，$y_{whistle}\ d\ GA_k$，$GA_k\ o\ y_{cheers}$，$y_{caption}\ e\ TV_{k+1}$，$y_{cheers}\ e\ TV_{k+1}$。如果上述所有对象都存在，则 $E_{Gk}$ 为一个进球事件。

其他事件也能通过相似的推理算法来进行探测，仅需要调整候选事件子集和要搜索对象的定义。

本节基于领域本体提出了基于 TDL 的事件建模方法和用于事件探测的推理算法。本文提出的事件描述和探测方法具有很强的灵活性和适应性，用户能够根据自己的领域知识定义和描述不同的事件并对相同事件采用不同的 TDL 描述。如对于进球事件，用户能够根据具体视频的领域知识定义一个与本文所用的不同的 TDL 描述。

## 6 实验与分析

本节利用足球领域的真实背景和相应的视频数据测试本文提出的框架。定义足球领域中的相关本体，采用 Protégé 本体构建工具创建本体，OWL DL 作为输出语言。

关于对象和序列探测的详细描述参见先前的工作[23,25]。本文主要讨论基于本体的视频内容分析框架的开发并论证所提出的基于 TDL 的事件建模方法和事件探测推理算法的有效性。为了避免自动探测带来的探测误差的影响，采用人工方法对足球视频中的对象和序列进行标注，得到视频的对象和序列数据集。

采集 5 场足球比赛视频进行实验。视频为 4∶2∶2 YUV PAL 制式、MPEG-1 格式，足球比赛的视频出自两个转播公司(ITV 和 BBC 体育)，均为 2006 世界杯比赛，总时长 7h53min28s。具体细节如表 2 所列。

表 2 足球比赛事件探测的实验数据

| 序号 | 比赛名称 | 转播公司 | 比赛时长 |
|---|---|---|---|
| 1 | 安哥拉 vs 葡萄牙 | ITV | 1:35:54 |
| 2 | 阿根廷 vs 塞黑 | BBC 体育 | 1:32:27 |
| 3 | 巴西 vs 法国 | ITV | 1:36:04 |
| 4 | 法国 vs 韩国 | BBC 体育 | 1:35:36 |
| 5 | 荷兰 vs 阿根廷 | ITV | 1:33:27 |

表 3 列出了语义事件探测的实验结果。其中“实际包含数目”为整个视频中人工识别的事件数目。从表 3 可以看出，事件探测结果的查准率高于 91%，而查全率的结果相对较低。这是因为 TDL 描述在逻辑上非常严格，不允许事件定义和探测到的事件之间出现任何细微的差别，因而推理算法对于事件探测能确保较高的查准率，但会丢失一些正确的结果。

表 3 3 种足球语义事件探测的实验结果

| 语义事件 | 实际包含数 | 正确探测数 | 错误探测数 | 查准率 % | 查全率 % |
|---|---|---|---|---|---|
| 进球 | 10 | 8 | 0 | 100 | 80 |
| 犯规 | 193 | 141 | 11 | 92.8 | 73.1 |
| 黄(红)牌 | 26 | 22 | 2 | 91.7 | 84.6 |

如果对同一事件定义不同的 TDL 描述，由不同的对象、序列和时域关系组成，将会获得高的查全率。例如，实验中丢失的进球事件具有不同的时域关系：GA-> OOF-> TV->MSR。给出附加的 TDL 定义如下：

$$Scoredgoal=\diamondsuit(d_{goal},d_{whistle},d_{cheers},d_{caption},d_{GA},d_{TV},d_{OOF},d_{SMR})(d_{goal}\ f\ d_{GA})(d_{whistle}\ d\ d_{GA})(d_{GA}\ o\ d_{cheers})(d_{caption}\ e\ d_{TV})(d_{cheers}\ e\ d_{TV})(d_{GA}\ m\ d_{OOF})(d_{OOF}\ m\ d_{TV})(d_{TV}\ m\ d_{MSR}).(goal@d_{goal}\cap whistle@d_{whistle}\cap cheers@d_{cheers}\cap caption@d_{caption}\cap GA@d_{GA}\cap TV@d_{TV}\cap OOF@d_{OOF}\cap SMR@d_{SMR})$$

在推理算法中采用上述定义能够探测到两个错失的进球事件。对同一事件采用不同描述的缺点在于增加了知识基础的复杂性。探测黄(红)牌事件错误的原因在于，在有些情况下，当一个球员被侵犯时，也出现了一个 MSR 和一个字幕对象。在这种情况下，黄牌事件的探测会出现错误。

基于上面的实验结果，可以相信本文提出的视频内容分析框架及基于 TDL 的事件描述和探测方法具有很大的潜力。下一步，将采用更大规模的数据集并将提出的框架应用到不同领域来进行验证，同时采用自动序列和对象探测标注结果进行试验。

**结束语** 本文提出了一个基于本体的视频语义内容分析框架。采用领域本体定义高层语义特征和它们在语义上下文中的联系，并将低层特征(如视觉和听觉特征)和视频内容分析算法集成进本体中，以增强本体语义表达与支持视频语义分析能力。

为了构建视频内容分析的领域本体，采用 OWL 本体描述语言，采用 DL 定义了如何根据不同的感知内容和低层特征选择视频分析的特征和算法的规则，采用 TDL 来描述语义事件。为了验证所提框架的有效性，采用 Protégé 构建一个足球领域本体进行实验。提出了基于 TDL 的事件建模方法和用于事件探测的推理算法。本文所提框架灵活、不依赖于视频低层分析任务并能应用在不同的领域中。实验显示所提框架对于视频语义层的内容分析是有效的。事件探测试验获得了较高的查准率而相对低的查全率，对此给出了分析解释。

未来的工作主要包括采用更复杂的模型表示和定义目标领域中复杂和重要的语义事件来增强领域本体，同时采用自

(下转第 178 页)

类算法，很大程度上克服了 K-Means 对初始代表点的依赖性，可以获得较高质量的聚类结果。但是，无疑 Meta-KMeans 算法的时间复杂度较 K-Means 算法要高很多，本文下一步工作将就该问题展开研究。

## 参 考 文 献


[1] Han J W, Kamber M. Data Mining: Concepts and Techniques. 2nd ed[M]. Morgan Kaufmann Publishers, 2001: 223-250

[2] Kirkpatrick S, Gelatt C D Jr, Vecchi M P. Optimization by Simulated Annealing[J]. Science, 1983, 220(4598): 671-680

[3] Glover F. Tabu search-Part I[J]. ORSA Journal on Computing, 1989, 1(3): 190-206

[4] Glover F. Tabu search-Part II[J]. ORSA Journal on Computing, 1990, 2(1): 4-32

[5] Dorigo M, Blum C. Ant colony optimization theory: a survey[J]. Theoretical Computer Science, 2005, 344(2/3): 243-278

[6] Boyan J A, Moore A W. Learning Evaluation Functions for Global Optimization and Boolean Satisfiability[C]// Jack Mostow, Chuck Rich, eds. Proc. of the 15th National Conference on Artificial Intelligence. CA, USA: AAAI Press, 1998

[7] Sebastiani F. A tutorial on automatic text categorization [C] // Anala Amandi, Ricardo Z, eds. Proc. of the 1st Argentinean Symposium on Artificial Intelligence. Buenos Aires, 1999: 7-35


(上接第 174 页)

动的低层特征分析和序列、对象探测；进一步探索分析多通道低层特征以抽取更准确、有效的语义内容表示；同时，本体的自动构建也是未来研究的一个重要方向。

## 参 考 文 献


[1] Chang S-F. The holy grail of content-based media analysis[J]. IEEE Multimedia, 2002, 9(2): 6-10

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

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

[4] Muller-Schneiders S, Jager T, Loos H S, et a. Performance evaluation of a real time video surveillance system[C]// 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. Oct. 2005: 137-143

[5] Hua X S, Lu L, Zhang H J. Automatic music video generation based on the temporal pattern analysis[C]// 12th Annual ACM International Conference on Multimedia. October 2004

[6] Informedia-II: Auto-Summarization and Visualization over Multiple Video Documents and Libraries[R]. September 2001. http://www.informedia.cs.cmu.edu

[7] Resource description framework. Technical report [EB/OL]. W3C. http://www.w3.org/RDF/, Feb. 2004

[8] Web ontology language (OWL) [EB/OL]. Technical report. W3C. http://www.w3.org/2004/OWL/, 2004

[9] Leonardi R, Migliorati P. Semantic index of multimedia documents[J]. IEEE Multimedia, 2002, 9(2): 44-51

[10] Ekin A, Tekalp A M, Mehrotra R. Automatic soccer video analysis and summarization[J]. IEEE Transactions on Image Processing, 2003, 12(7): 796-807

[11] Yu X, Xu C, Leung H, et al. Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video[C]// ACM Multimedia 2003. Berkeley, CA (USA), Nov. 2003, 3: 11-20

[12] Xu H X, Chua T-S. Fusion of AV features and external information sources for event detection in team sports video[J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2006, 2(1): 44-67

[13] Reidsma D, Kuper J, Declerck T, et al. Cross document ontology based information extraction for multimedia retrieval[C]// Supplementary Proceedings of the ICCS03. Dresden, July 2003

[14] Mezaris V, Kompatsiaris I, Boulgouris N, et al. Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(5): 606-621

[15] Jaimes A, Tseng B, Smith J. Modal keywords, ontologies, and reasoning for video understanding[C]// International Conference on Image and Video Retrieval (CIVR 2003). July 2003

[16] Jaimes A, Smith J. Semi-automatic, data-driven construction of multimedia ontologies[C]// Proc. of IEEE Int'l Conference on Multimedia & Expo. 2003

[17] Bertini M, DelBimbo A, Torniai C. Enhanced ontoloies for video annotation and retrieval[C]// ACM MIR'2005. Singapore, November 2005

[18] Sadlier D A, O'Connor N E. Event Detection in Field Sports Video Using Audio-visual Features and A SVM [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(10): 1225-1233

[19] Dasiopoulou S, Papastathis V K, Mezaris V, et al. An Ontology Framework for Knowledge-Assisted Semantic Video Analysis and Annotation [C]// Proc. 4th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2004) at the 3rd International Semantic Web Conference (ISWC 2004). November 2004

[20] Strintzis J, Bloehdorn S, Handschuh S, et al. Knowledge representation for semantic multimedia content analysis and reasoning[C]// European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology. Nov. 2004

[21] Kompatsiaris I, Mezaris V, Strintzis M G. Multimedia content indexing and retrieval using an object ontology[A]// Stamou G, ed. Multimedia Content and Semantic Web Methods, Standards and Tools. New York: Wiley, 2004

[22] Artale A, Franconi E. A temporal description logic for reasoning about actions and plans[J]. Journal of Artificial Intelligence Research, 1998, 9: 463-506

[23] Chen J Y, Li Y H, Lao S Y, et al, Detection of Scoring Event in Soccer Video for Highlight Generation[R]. National University of Defense Technology, 2004

[24] Pan Hao, van Beek P, Sezan M I. Detection of Slowmotion Replay Segments in Sports Video for Highlights Generation[C]// Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP'01). Salt Lake City, UT, USA, May 2001

[25] Bai Liang, Hu Yanli, Lao Songyang, et al. Feature Analysis and Extraction for Audio Automatic Classification[C]// IEEE SMC 2005. Hawaii USA, October 2005

[26] Zhou W, Dao S, Jay Kuo C-C. On-line knowledge and rule-based video classification system for video indexing and dissemination [J]. Information Systems, 2002, 27(8): 559-586

[27] MPEG-7 Overview[OL]. http://www.chiariglione.org, October 2004