

文章编号: 1001-2486(2009)05-0050-04

# 一种新的足球视频语义镜头表示及事件查询框架\*

卜江, 老松杨, 刘海涛, 郭金林

(国防科技大学 信息系统与管理学院, 湖南 长沙 410073)

**摘要:** 提出了一个足球视频中的语义镜头表示及事件查询框架。在这个框架中, 定义了不同的语义镜头, 并且每类语义镜头都可以由五个特征属性所刻画。在语义镜头定义的基础上, 为了弥补传统的事件查询方式的不足, 提出了一种用户定义事件的查询方式, 它是一种基于语义镜头合成 Petri 网表示模型的查询方式, 可以由用户自由定义所感兴趣的事件 Petri 网表示结构, 实验证明该查询方式能有效提高查询效率, 并有很大的灵活性, 能适应不同广播公司的转播规则, 但该查询方式需要用户具有一定的专业和领域知识。实验数据来源于 2006 年德国世界杯、英格兰足球超级联赛和西班牙足球甲级联赛。

**关键词:** 足球视频; 语义镜头表示; 合成 Petri 网; 事件查询

中图分类号: TP37 文献标识码: A

## A New Framework for Semantic Shot Representation and Event Query of Soccer Video

BU Jiang, LAO Song-yang, LIU Hai-tao, GUO Jin-lin

(College of Information System and Management, National Univ. of Defense Technology, Changsha 410073, China)

**Abstract:** In this paper, a semantic shot representation and event query framework for soccer video is proposed. In this framework, various semantic shots are defined and each shot can be described by five properties. In order to solve the problem in traditional event query, a user-defined event query method is also proposed, which utilizes the Semantic Shot Composite Petri-Nets. Based on the representation of the semantic shot, users can define their favorite events by Petri-Nets. The experiment conducted shows that the query method can greatly improve the query efficiency and adjust self-adaptive as the broadcaster varies, but this method requires users to have enough professional knowledge. Experiment data set includes Germany World Cup 2006, English Premier League and Spanish First Division.

**Key words:** soccer video; semantic shot representation; composite Petri-nets; event query

多媒体信息数量的飞速增长使得基于内容的视频索引、检索和数据管理技术迅速发展<sup>[1-2]</sup>。在足球视频检索研究方面, 目前已有的检索方式多为传统的文本语义描述方式, 即输入某个关键字来查找相应的视频片段。在这种查询方式下, 用户查询到的感兴趣事件在事件探测算法中体现为固定的语义库和语义规则, 其缺乏灵活性, 由于不同的广播公司可能存在不同的转播规则, 传统的查询方式就会导致较大的误差。文献[7]在用户使用户能够干预和调整语义规则的制定以及个性化的查询定义这个问题上做出了有益的尝试, 它通过采用基本语义单元合成 Petri 网(Basic Semantic Unit Composite Petri-net, BSU CPN)模型对足球视频中感兴趣事件进行了基于语义单元表示, 但是它缺乏对底层镜头特征的统一表示, 通用性和扩展性较差。本文在此基础上, 将所提出的语义镜头表示框架和语义镜头的合成 Petri 网查询描述模型统一起来, 建立了用户定义事件查询方式, 从而很好地解决了传统查询方式存在的问题, 最后的实验结果证明其具有较好的灵活性和准确性。

### 1 足球视频语义镜头表示框架

提出的模型结构如图 1 所示。在对足球视频进行镜头探测后, 所得到的单个镜头可由以下五个特征属性进行描述: 镜头类型、镜头对象、镜头场地区域、镜头人物、镜头音频特征, 它基本上覆盖了足球视

\* 收稿日期: 2009-03-16

基金项目: 国家 863 计划资助项目(2006AA01Z316); 国家自然科学基金资助项目(60572137)

作者简介: 卜江(1983-), 男, 博士生。

频中所出现的所有镜头类型。根据这五个特征,可以给这些镜头赋予一定的语义,语义镜头就定义为其特征元素不全为空值的镜头。

一个语义镜头 = { 镜头类型, 镜头对象, 镜头场地区域, 镜头人物, 镜头音频特征 }

在足球视频中,镜头类型集包括远景镜头、中景镜头、特写镜头、慢镜头;镜头对象集包括字幕对象、球门对象、角弧对象;镜头场地区域集包括左上半场区域、左下半场区域、上中场区域、下中场区域、右上半场区域、右下半场区域;镜头人物集包括球员、裁判、教练、观众;镜头音频特征集包括哨声、欢呼声。

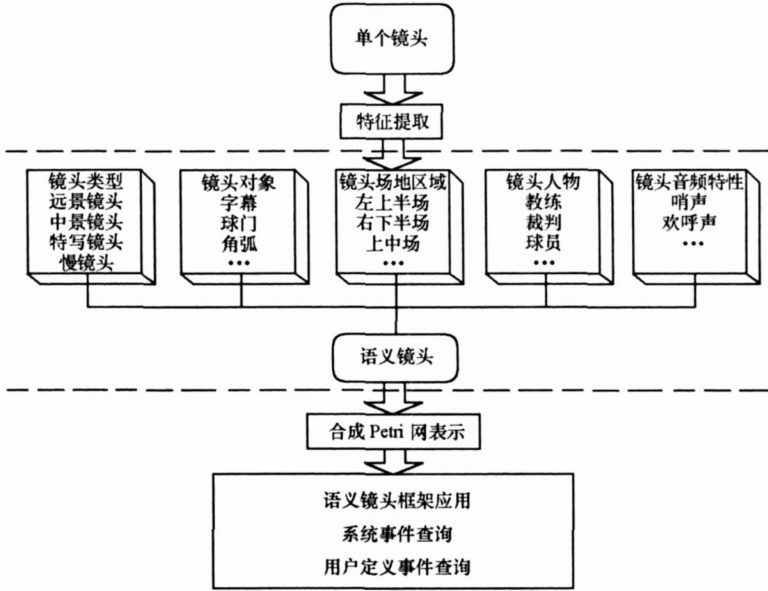


图 1 足球视频语义镜头表示框架图

Fig. 1 The framework of semantic shot representation in soccer video

基于所提出的足球视频语义镜头表示框架,可以利用语义镜头进行基于事件的语义分析以及系统定义事件和用户定义事件的查询。

## 2 语义镜头定义

通过上节所定义的五個特征,我们可以定义足球视频中的语义镜头。在这里截取了 2006 年世界杯足球比赛视频的一些图片作为语义镜头的实例,如图 2 所示。

通过联合这五个特征形成五维特征向量来定义一个语义镜头:

$$semanticshot_i = \{ st_i, so_i, fa_i, sc_i, sf_i \} \text{ 且 } st_i \in ST, so_i \in SO, fa_i \in SFA, sc_i \in SC, sf_i \in SAF \quad (1)$$

其中,  $semanticshot_i$  表示第  $i$  个语义镜头,  $st_i, so_i, fa_i, sc_i, sf_i$  分别表示第  $i$  个语义镜头的镜头类型, 镜头对象, 镜头场地区域, 镜头人物和镜头音频特征,  $ST, SO, SFA, SC, SAF$  分别为镜头类型集、镜头对象集、镜头场地区域集、镜头人物集和镜头音频特征集。

## 3 基于语义镜头的事件查询方式

利用文献[7]中的建模表示思想,提出了基于语义镜头的事件查询描述模型,在此模型基础上,定义了两种用户查询方式:系统定义事件查询和用户定义事件查询。

(1) 基于语义镜头的合成 Petri 网模型(SSCPN)

主要有两种表示形式,一种是基于语义镜头特征属性的表示模型,另一种是相应的语义镜头表示模型。

SSCPN 是一个七元式,定义如下:

$$C_{SSCPN} = \{ P, T, F, Du, SS, SSF, Op \} \quad (2)$$

其中,  $P = \{ p_1, p_2, \dots, p_m \}; T = \{ t_1, t_2, \dots, t_n \};$

$F \subseteq \{ P \times T \} \cup \{ T \times P \}; Du: P^{\rightarrow} (Operate, Duration), Operate = \{ \leq, =, \geq \}, Duration \in R;$



图 2 语义镜头的实例图

Fig. 2 The instance of semantic shot

$SS: P \rightarrow \{null, SemanticShot\}$ ;  $SSF: P \rightarrow \{null, SemanticShotFeature\}$ ;  
 $Op: T \rightarrow \{and, or, following, before, synchronization\}$ .

这里,  $P, T, F, Du, SS, SSF, Op$  分别代表库所集合, 变迁集合, 有向弧集合, 库所的持续时间, 语义镜头集合, 语义镜头特征集合, 变迁的点火逻辑。其中  $Op$  包括与或逻辑、前后顺序和时间同步三类。可以利用所建立的语义镜头合成 Petri 网模型(SSCPN)对足球视频中的重要事件进行建模, 如进球、犯规、换人、角球等。采用两种形式表示的主要目的是将语义镜头特征与提出的语义镜头表示框架联系起来, 由于语义镜头表示框架可以很容易扩展到其他球类视频中, 从而使得基于语义镜头的合成 Petri 网模型会有较好的可扩展性。给出足球视频中进球这个语义事件的建模示例, 如图 3、图 4 所示。其中, 图 3 利用了语义镜头特征集进行进球事件表示, 图 4 则利用了语义镜头集进行进球事件表示。

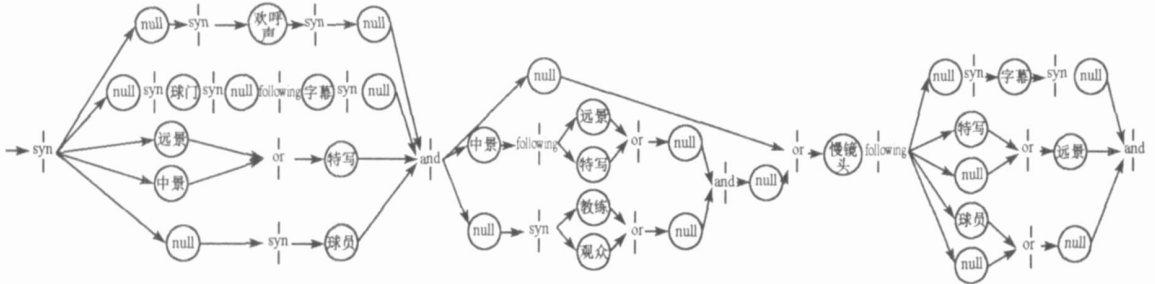


图 3 进球事件的语义镜头合成 Petri 网模型表示一

Fig. 3 Goal events described with semantic shot using Petri nets model iv

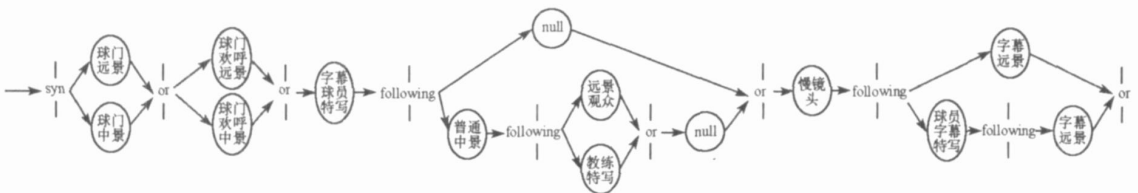


图 4 进球事件的语义镜头合成 Petri 网模型表示二

Fig. 4 Goal events described with semantic shot using Petri nets model ⑤

(2) 系统定义事件与用户定义事件查询

在上节语义镜头合成 Petri 网基础上可以实现系统定义事件查询。所谓系统定义事件查询是指由

体育视频分析和检索系统通过已给定的事件 Petri 网描述结构,自动探测得到体育视频中的重要事件,并提供给用户进行基于事件关键字的查询。这种查询方式是一种较常见的方式,传统的视频摘要和检索系统一般是基于这种查询方式,只需观察学习体育视频中典型事件的规律就可以较容易得到,但该查询方式存在扩展性和灵活性较差的缺点<sup>[7]</sup>。

为了解决系统定义事件查询存在的缺陷,提出了一种新的用户定义事件查询方式。所谓用户定义事件查询,指的是用户可以通过系统提供的语义镜头集、语义镜头特征集、变迁点火逻辑集等,利用合成 Petri 网,结合用户自身对事件的理解,由用户给出感兴趣事件的语义 Petri 网描述,系统再利用这个描述结构进行事件探测的查询方式。

用户定义事件查询方式的具体流程为:

- (1) 用户从语义镜头特征集、语义镜头集、变迁逻辑集中拖取相应的图标至事件的 Petri 网表示栏;
- (2) 用户使用有向弧将所拖取出的元素按照自己所定义的 Petri 网结构连接起来;
- (3) 如需要定义库所持续时间  $Du$ ,则在相应元素位置输入持续时间值;
- (4) 提交给查询系统进入数据库进行所定义的事件查询。

这种查询方式能较好地解决传统查询方式存在的问题。首先事件的 Petri 网表示由用户给出,因此用户可以根据自己的需求定义自己所谓的感兴趣事件;其次用户在定义 Petri 网表示结构时,它一定对该视频的广播公司的转播规则有一定的了解,对于不同广播公司的视频会设定不同的结构,因此可以大大降低转播规则的不同所带来的查询误差,并且具有很好的灵活性。

## 4 实验结果

为了验证语义镜头表示及查询框架的有效性,选择了 2006 年德国世界杯、英格兰超级联赛和西班牙足球甲级联赛的比赛片断进行实验,如表 1 所示。在经过足球视频分析处理后,进行了两类查询,查询结果如表 2、表 3 所示。

表 1 足球视频实验数据

Tab. 1 Experimental data of soccer video

视频 ID	比赛名称	广播公司	比赛队伍	比赛时间	视频持续时间
1	德国世界杯	CCTV5	巴西 vs. 加纳	2006 年	17m10s
2	德国世界杯	CCTV5	意大利 vs. 德国	2006 年	1h25m46s
3	英格兰超级联赛	湖南卫视	曼联 vs. 利兹联	2003 年	8m3s
4	西班牙甲级联赛	CCTV5	巴萨 vs. 皇马	2007 年	55m14s

表 2 系统定义事件查询结果

Tab. 2 Query results of system-defined events

足球事件	实际数	探测数	正确数	查准率	查全率
进球	9	10	8	80%	89%
换人	12	14	10	71%	83%
角球	28	30	25	83%	89%
犯规	53	60	45	75%	85%
进攻	142	159	113	71%	80%
平均值				76%	85.2%

表 3 用户定义事件查询结果

Tab. 3 Query results of user-defined events

足球事件	实际数	探测数	正确数	查准率	查全率
进球	9	9	8	89%	89%
换人	12	13	11	85%	92%
角球	28	30	27	90%	96%
犯规	53	59	49	83%	92%
进攻	142	156	131	84%	92%
平均值				86.2%	92.2%

- [6] 张逢星,魏小兰,史启祯,等.三元体系  $MgCl_2 - CO(NH_2)_2 - H_2O$  在 25℃ 时的等温溶解度与新相研究[J]. 高等学校化学学报, 1998, 19(9): 1449- 1452.
- [7] 杨吉民,陈见见,张润芝. 20℃ 时  $NaCl - C_2H_5OH - H_2O$  三元体系的相平衡研究[J]. 河南化工, 2007, 24(2): 20- 22.
- [8] 叶明泉,韩爱军,马振叶,等. 超细粒子及其复合技术在冷气溶胶灭火剂中的应用[J]. 南京理工大学学报, 2005, 29(2): 236- 239.
- [9] 胡维杰,宁桂玲. 结晶过程[M]. 大连: 大连理工大学出版社, 1991: 21- 49.
- [10] 宋海香,孙保平,茹宗玲. 防结块剂的研究方法和进展[J]. 化工进展, 2001(6): 50- 52.
- [11] Hansson H C. NaCl Aerosol Particle Hygroscopicity Dependence on Mixing with Organic Compounds[J]. Journal of Atmospheric Chemistry, 1998(31): 321- 346.
- [12] 周甜,钱刚,周兴贵,等. 超声波对甘氨酸溶析结晶过程的影响[J]. 过程工程学报, 2007, 7(4): 728- 732.
- [13] 田军. 超声波在碱式氯化镁结晶中的应用[J]. 哈尔滨商业大学学报, 2005, 21(2): 221- 222.
- [14] 李先华,党乐平,殷萍. 硫酸铵蒸发结晶过程影响因素研究[J]. 无机盐工业, 2008, 40(11): 40- 43.
- [15] 党亚固,胡秀英,费德君,等. 磷酸二氢钾结晶的影响因素研究[J]. 无机盐工业, 2008, 40(6): 16- 18.
- [16] 郁平,朱贤,陶建伟. 多孔球形超细碳酸钙制备及其生成机理[J]. 上海交通大学学报, 2006, 40(11): 1979- 1982.
- [17] 朱裕贞,顾达,黑恩成. 现代基础化学[M]. 北京: 化学工业出版社, 1998: 90- 92.
- [18] 白世河,潘伟,房明浩,等. 硬脂酸在制备超细  $\alpha - Al_2O_3$  粉体中的防团聚作用[J]. 稀有金属材料与工程, 2005, 34(1): 164- 166.
- [19] 胡维杰,宁桂玲. 结晶过程[M]. 大连: 大连理工大学出版社, 1991: 21- 49.
- [20] Zhao D Y, Huo Q S, Feng J L, et al. Nonionic Triblock and Star Diblock Copolymer and Oligomeric Surfactant Syntheses of Highly Ordered, Hydrothermally Stable, Mesoporous Silica Structures[J]. J. Am. Chem. Soc, 1998, 120: 6024- 6036.
- [21] Gulun M A, Nguyen M H, Kriven W M. Polymerized Organic-inorganic Synthesis of Mixed Oxides. Cream[J]. J. Am. Chem. Soc, 1999, 121: 556- 560.
- [22] 姚连增. 晶体生长基础[M]. 合肥: 中国科技大学出版社, 1995: 135- 146.
- [23] McCausland L, Cains P. Ultrasound to Make Crystals[J]. Chem. Ind., 2003(5): 15- 17.
- [24] Guo Z, Zhang M, Li H, et al. Effect of Ultrasound on Anti-solvent Crystallization Process[J]. J. Cryst. Growth, 2005, 273: 555- 563.

(上接第 53 页)

从实验结果整体来看,事件查询结果的查全率和查准率总体都较低,主要原因是事件探测本身存在误差,而在进行底层语义镜头探测和语义镜头特征提取时也会存在较大误差,因此改进底层特征提取和语义镜头探测算法的效率是需要解决的一个问题。对比以上两类查询方式的结果可以看到,用户定义事件查询结果在一定程度上有所提高,这主要是由于实验数据集包含了两个广播公司的视频,两公司对于比赛的转播规则存在差异,采用用户定义事件查询时,用户可以根据不同的广播公司对同一类事件发生的规律进行总结分析,设计不同的 Petri 网结构进行事件表示,即 Petri 网结构可以根据广播公司的不同而进行自适应的调整,从而改善事件查询结果的精确度。

## 5 总结与展望

本文提出了一个新的足球视频语义镜头表示及事件查询框架。结合足球领域知识,根据特征属性对语义镜头进行细致分类。在此基础上,基于语义镜头 Petri 网表示模型,提出了一种用户自定义语义事件的查询方式,以弥补传统语义事件查询方式的不足。实验证明该查询方式灵活有效,能适应不同广播公司的转播规则。

实验结果显示查询结果总体还是偏低,其中底层特征提取算法和语义镜头探测算法的改进是今后需要解决的一个问题。另外,如何将这种表示查询框架推广到其他球类视频,如篮球、网球等,也是值得研究的一个内容。

## 参考文献:

- [1] Smeulders A W M, Worring M, Gupta A, et al. Content Based Image Retrieval at the End of the Early Years[J]. IEEE Trans. PAMI., 2000, 22(12): 1349- 1380, 2000.
- [2] Alejandro J, Christel M, Sebastien G, et al. Multimedia Information Retrieval: What is it, and Why isn't Anyone Using It? [C]//Proc. of ACM MIR, 2005: 3- 8.
- [3] Xu C S, Wang J J, Wan K W, et al. Live Sports Event Detection Based on Broadcast Video and Web-casting Text [C]//Proc. of ACM Multimedia, 2006: 221- 230.
- [4] Bezerra F N, Lima E. Low Cost Soccer Video Summaries Based on Visual Rhythm[C]//Proc of ACM Multimedia Information Retrieval, 2006: 71- 77.
- [5] Snoek C, Worring M. Multimedia Event Based Video Indexing using Time Intervals [J]. IEEE Trans. Multimedia, 2005.
- [6] Tong X F, Liu Q S, Duan L Y, et al. A Unified Framework for Semantic Shot Representation of Sports Video [C]//Proc. of ACM Multimedia, 2005: 127- 134.
- [7] Lao S Y, Smeaton A F, Jones G J F, et al. A Query Description Model Based on Basic Semantic Unit Composite Petri-nets for Soccer Video Analysis [C]//Proc. of ACM Multimedia, 2005: 143- 150.