

一种鲁棒的场地球类视频中的球门探测方法*

卜江, 郑彬, 廖东升

(国防科技大学人文与社会科学学院, 湖南长沙 410074)

摘要:在场地球类视频中,球门常常会伴随着精彩片段而出现。因此,球门探测是体育视频摘要的基础,同时也是视频高层语义概念探测的研究热点。目前,利用机器学习方法进行视频中的对象探测是非常有前途的研究领域。基于此,提出一种基于模糊决策树的球门探测算法,用来探测场地球类视频中球门帧的出现,为了提高分类准确性,在模糊决策树训练的过程中加入了平衡处理。实验结果表明,与基于阈值和决策树的算法相比,该算法可以得到更好的分类结果(F-measure > 95%),并且可以从所建立的树中推导出模糊规则来解释分类模型。

关键词:球门探测;模糊决策树;场地球类视频;平衡处理

中图分类号:TP37 **文献标志码:**A **文章编号:**1001-2486(2013)05-0085-06

A robust goalmouth detection method in field-ball game video

BU Jiang, ZHENG Bin, LIAO Dongsheng

(College of Humanities and Social Sciences, National University of Defense Technology, Changsha 410073, China)

Abstract: Recently, goalmouth detection has become the basis of highlight summarization and a hot topic in high-level semantic concept detection, as the appearance of goalmouth indicates a high likelihood of exciting action in field-ball game video. In addition, machine learning for object detection is a promising research frontier. A robust method is proposed to detect the presence of goalmouth frame in field-ball game video based on fuzzy decision trees. Balance process was added in training procedure. The experimental results show that our algorithm (F-measure > 95%) can improve the classification performance in comparison with the threshold-based algorithm and the decision tree-based algorithm. Fuzzy rules can also be easily deduced from the constructed tree to interpret the classification model.

Key words: goalmouth detection; fuzzy decision tree; field-ball game video; balance process

在高速宽带网络和数字视频技术发展的今天,用户有了大量的视频数据可供选择。与此同时,体育比赛在全球备受观众青睐,并且体育视频转播在公共和商业广播中占据了很大的份额。这种情况引发了一些极具挑战的研究问题:如何根据用户的不同喜好来快速找到感兴趣的体育视频片断,另外一个紧迫的需求是如何利用增强和表示技术给观众提供更好的观看效果,这需要在虚拟现实和增强现实领域进行相关的研究^[1]。

概念探测是一个跨越“语义鸿沟”的有效方法,而球门是体育视频中的重要对象并且球门探测可以看成是体育视频中的一种高层语义概念探测方法。一旦探测到了球门,下一步的工作可以是推断不同的体育比赛高层语义事件或者在球门周围插入广告和虚拟内容。在已有的研究工作中,文献[2]中的 Kongwah 等通过定位球门的像素坐标来探测球门,他们还在文献[3]中提出了一种在 MPEG 足球视频中实时探测球门的方法,

实验中虽然取得了比较好的查全率和查准率,但是该算法仅能在足球视频中进行球门探测并且需要预先定义很多阈值。

本文主要提出了一种鲁棒的基于模糊决策树的球门探测方法,该算法可以在场地球类视频中探测球门帧的出现。由于数据挖掘方法能很好地建立球门出现时所具有的相同的内在模式,因而它有助于体育视频内容结构的分析,并且数据挖掘方法在数字视频处理中的应用是当前比较热门的研究领域。已有文献中相关的研究工作包括 Marsala 等提出一种模糊数据挖掘方法来处理视频^[4],并且他们在 TRECVID2008 的对象探测任务中,利用模糊决策树森林来探测不同的高层特征^[10],Fan 等使用支撑向量机来探测监督视频中的人物^[5],Zhao 等利用 AP(Affinity Propagation)聚类算法来探测图像中的对象^[6-7],Liang 等利用决策树来对棒球视频进行探测和摘要^[8],Zhang 等提出了一种基于主成分分析的算法来探测视频中的运动对象^[9],文献

* 收稿日期:2013-03-28

基金项目:国防科技大学校预研项目(JZ12-08-08);国家部委资助项目;国家自然科学基金资助项目

作者简介:卜江(1983—),男,广东增城人,讲师,博士,E-mail: veron9@163.com

[11]中的作者提出了基于学习的方法来探测足球视频中感兴趣的事件,文献[12]利用 Top - Hat 变换和一系列推导式的规则进行球门探测,并最终用于精彩事件的提取。该算法的不足之处是实验数据中仅包含了足球视频,并且实验结果中的查全率(88.94%)和查准率(86.18%)还有待提高,同时该算法的推导规则需要预先确定大量的阈值。

本文的贡献主要在于:

(1) 提出的基于模糊决策树的方法可以在不同的体育类型,比赛类型和比赛条件下探测场地球类视频中的球门;

(2) 在模型训练阶段增加了平衡处理(Balance Process),使得训练数据中的不同类别(Positive 和 Negative)的数量是相等的,以此来提高分类的准确性;

(3) 与基于阈值和基于决策树的方法相比,本文方法有效改进了球门探测的精确度;

(4) 根据所构建的模糊决策树模型,可以推导相应的模糊规则进行解释。

1 场地球类视频特点

由于体育比赛都具有特定的比赛规则,而且大部分的体育比赛都是在有范围限制的比赛场地中进行,因此与别的视频类型相比(如家庭拍摄视频),体育比赛视频具有更加固定的结构。如果能够利用这些特定的领域知识来进行体育视频分析,将会得到更好的语义层次的结果。

定义场地球类比赛是由许多不同的体育比赛所组成的集合,它们通常由两支队伍在有范围限制并且包含两个球门的场地中进行比赛。

一般来说,我们只关注场地比赛视频的一些独有的特征,主要包括:

(1) 场地球类比赛中的镜头类型与其他视频源的镜头类型不同(如新闻视频和电影等),在场地球类视频中我们仅仅观察到三种清晰的镜头类型:远景镜头、中景镜头和特写镜头。

(2) 在场地球类视频中,“远景镜头”里球门的出现是非常重要的,因为它很有可能意味着一些重要的语义事件,例如进球,进攻等。

(3) 球门通常包含两个垂直的门柱和一个水平的横梁,它们具有特定的颜色,并且“远景镜头”中球门的出现一般伴随着高能量语音。关于场地球类视频中球门的不同类型,见图 1。

2 模糊决策树

决策树是一种类似于流程图的结构,其中每个

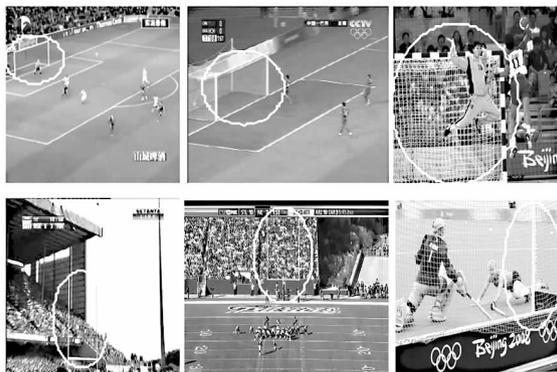


图 1 场地球类视频中的各类球门帧

Fig. 1 Different appearance of goalmouth in field-ball game video

节点代表对某个属性的测试,每一个分支代表测试的输出,叶节点代表对于实例的分类。通常的决策树分类是从顶部进行到底部。每个节点上所提出的问题意味着模式的一个特殊的属性,而向下的连接意味着可能的值。通过访问连续的节点直至终点或者叶节点,在这里可以读取类别标签。

归纳式的学习过程一般经历由特殊到一般的过程,已有的研究工作表明,归纳式学习方法在处理符号型数据时效果很好,但在处理数值型或数值-符号型的数据时就容易出现问題,而模糊集理论(Fuzzy Set Theory)定义了处理数值/符号型数据的有效工具,由于本文中所提取的场地球类视频特征为数值型数据,因此考虑使用模糊决策树进行分类模型的构建。

模糊决策树在以下两个方面不同于传统的决策树:首先它使用基于模糊约束的划分准则,并且它的推理程序是不同的。算法被强加了模糊集所定义的建立树的模糊条件。其次,在模糊决策树中,节点与数值和符号型属性相关联,边由模糊态所标注,叶子则能够与类的模糊子集相关联。通过视频数据来产生模糊规则的算法流程图如图 2 所示。

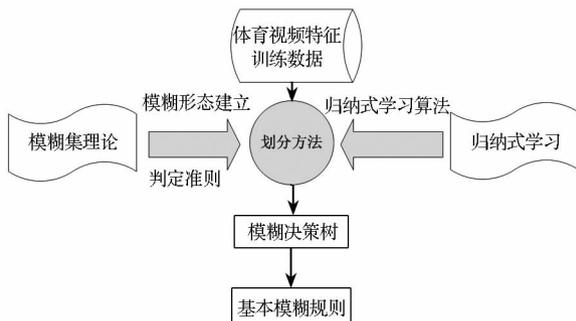


图 2 基于体育视频数据的模糊规则产生

Fig. 2 Generation of fuzzy rule

建立模糊决策树的方法根据其所选择的判定

准则不同而不同。在处理数值/符号型数据时,通常使用的判定准则是测量模糊事件的熵,叫做星熵(Star-Entropy)。星熵的公式如下:

$$H_S^*(C | A_j) = - \sum_{l=1}^{m_j} P^*(V_{jl}) \sum_{k=1}^K P^*(C_k | V_{jl}) \log(P^*(C_k | V_{jl})) \quad (*)$$

在(*)中,类C的集合表示一个物理或者概念上的现象,属性集 $A = \{A_1, \dots, A_N\}$ 可以用来描述这种现象。每个属性 A_j 可以在 $\{v_{j1}, \dots, v_{jm_j}\}$ 中取值 v_{jl} 。所谓一个描述(Description)指的是一个有关于属性值对 (A_j, v_{jl}) 的N元式。每个描述与一个 $C = \{c_1, \dots, c_K\}$ 中特殊的类 c_k 相关联,并以此构造现象的一个实例 e_i 。星熵准则是从经典的香农熵(Shannon Entropy)变换而得到,香农熵中的经典概率被Zadeh所提出的模糊事件的概率值 P^* 替换,从而成为星熵^[14]。

本文实验部分的模糊决策树构造和分类工作主要由本实验室所开发的Salammbô软件来实现,而星熵也是Salammbô软件中所实现的一种判定准则^[13]。Salammbô软件通过实现一个分层次的FGH模型来建立判别准则(见图3)^[14],该准则可以在归纳学习过程中评价用来测量属性判别能力的函数的有效性。

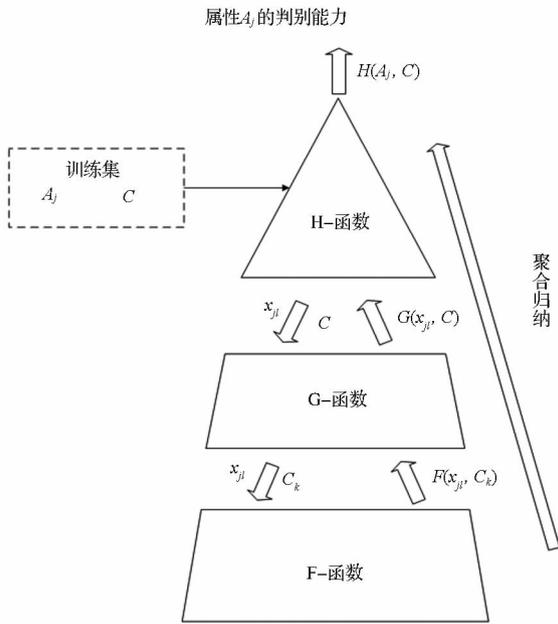


图 3 函数的分等级模型表示

Fig. 3 Hierarchical models of functions

3 基于模糊决策树的球门探测

监督式学习是一种机器学习方法,它能够从训练数据中学习得到映射函数,并且能够生成从

输入对象到分类结果的全局模型。本文中,我们尝试使用模糊决策树中来完成探测工作。主要进行以下三个步骤:

- (1) 通过平衡处理建立训练数据集;
- (2) 确定输入特征的表示形式;
- (3) 通过训练得到分类模型后,在视频流中进行球门探测。

3.1 通过平衡处理建立训练数据集

经过压缩编码的视频流可以包含大量的视频帧,例如,每秒视频流中可编码25~30个视频帧,那么,一个小时的视频流中将会有接近一百万左右的视频帧,如果在所有的视频帧中都运行球门探测算法,计算量将会非常巨大。因此,结合第1部分所分析的场地球类视频的特征,可以将视频流中的球门探测工作转换成关键帧的分类工作。

事实上,由于决策树的构造方法基于这样一个假设,即各个类别的数量是平均分布的,因此,为了得到一个有效的训练集来构造模糊决策树,在训练前需要通过平衡处理来保证每个类别中关键帧的数量是相等的。同时,为了通过分类得到“远景镜头”中的球门帧,实验中仅仅把“远景镜头”中的球门帧作为肯定类别(Positive)的训练数据,由于在否定类别(Negative)的训练数据中包含了三类镜头(即远景镜头、中景镜头和特写镜头),所以为了完全地表示否定类别的训练数据,同样也有必要平衡这三类镜头帧。

为了平衡训练数据,两个预先的处理将要在特征提取前实施,即镜头探测和镜头分类。在已有的文献中已经有许多有效的算法,本文将采用我们在文献[15-16]中所提出的算法进行镜头探测和分类。在进行镜头探测和镜头分类之后,我们从三类镜头中随机选取相同数量的否定类别(Negative)关键帧,并且从“远景镜头”中随机选取与前面三种否定类别(Negative)关键帧数量之和相等的肯定类别(Positive)关键帧,从而得到最终的训练数据。

3.2 特征表示

由于在场地球类视频中,“远景镜头”里球门的出现包含了共同的特定的特征,如视觉特征和听觉特征等,因此我们主要考虑以下几个特征描述符:

- 视觉特征描述符(见图4)
- 颜色特征

颜色提供了一种计算快速并且可靠的特征来探测潜在的目标区域。由于光照条件、天气状况、拍摄角度等因素的变化,球门通常呈现不同的外

观和颜色(即在白色值周围波动)。为了在视频流中探测球门的出现,本文利用两个颜色空间进行球门的颜色特征表示,即在 HSV(色调,饱和度和亮度)空间和 RGB(红色,绿色和蓝色)空间中提取整个关键帧和其子区域的特征值分量,从而充分利用不同颜色空间的互补性。

• 边缘特征

边缘一般用来分离图像中不同的内容,它在颜色或者亮度有变化的情况下是鲁棒的。由于体育视频中的球门在任何拍摄角度下都是竖直的,因此,本文仅考虑垂直方向上的边缘点,即在整个关键帧和其子区域内提取利用垂直方向的算子所得到的 Sobel 边缘点。

• Hough 变换线

由足球比赛的规则可知,球门由一个 7.32m 的横梁和两个 2.44m 的立柱构成,这将导致其在视频帧中的长度是固定的。因此,本文在整个关键帧内提取具有一定长度的 Hough 变换线条。

• SIFT 特征

为了使所提取的球门特征点对于图像尺度和旋转不变,同时对于视角和亮度变化是鲁棒的,我们在整个关键帧内提取 SIFT 特征点。

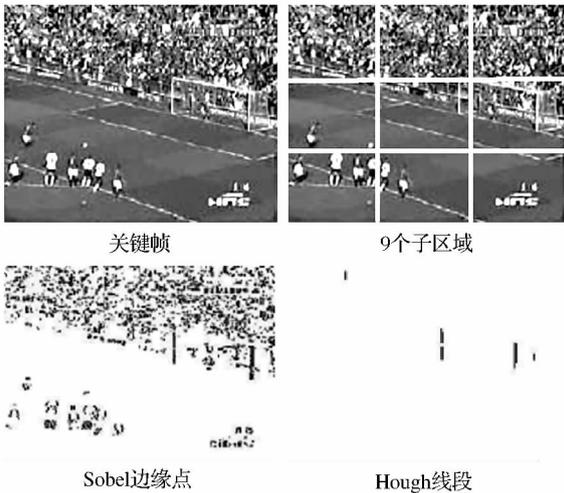


图 4 不同的视觉特征描述符

Fig. 4 Different visual feature descriptors

• 听觉特征描述符

在体育比赛中,球门与精彩事件(如进球,射门、扑救等)的联系非常紧密,而精彩事件的出现必然伴随着解说员和观众的欢呼。本文从每一个音频片段中提取 MFCC 特征,该特征能够表明视频流中的高能量语音。

• 空间结构描述符

图像帧的空间结构即为图像帧中各个单独元素的布局。由于“远景镜头”一般由架设在主席台

附近的主摄像机所拍摄,并且其与两侧球门区域具有特定的角度,因此“远景镜头”图像帧的空间结构与球门也必然紧密相连。本文利用 9 个块状子区域来进行颜色和边缘特征的提取,如图 4 所示。

最后,为了降低所提取的数值特征的维数,我们仅用每个区域特征的中值和方差作为最终的特征描述符。

3.3 视频流中的球门探测

在得到基于模糊决策树的分类模型后,我们使用它来判定视频流中的球门是否出现。根据我们的观察,“远景镜头”中球门的出现时间一般长于 2s,因此我们将测试的模式设置为在连续的视频流中每间隔 2s 提取一个关键帧进行探测,并且所提取的关键帧的位置取决于帧率,然后再从每个测试关键帧中提取相应的数值特征。

图 5 所表示的是球门探测算法的流程图,主要包括预处理、训练和测试三个过程。预处理过程

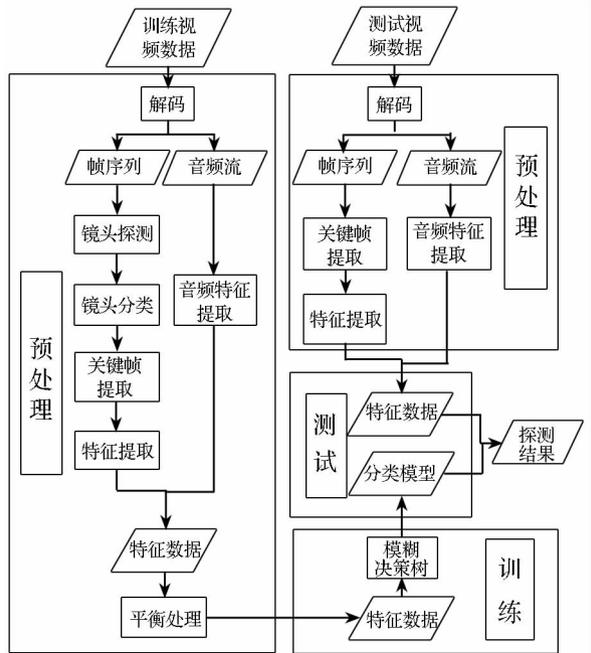


图 5 基于模糊决策树的球门探测算法流程图

Fig. 5 The flowchart of proposed algorithm

可以从原始视频数据(包括训练数据和测试数据)中提取数值特征,训练过程则利用训练数据的数值特征归纳得到分类模型,而测试过程通过所得到的模型来对测试数据进行分类以得到分类结果,即球门探测结果。

4 实验

4.1 数据准备

选取了七场具有不同压缩格式(MPEG, AVI 和 RMVB),不同体育类型(足球,橄榄球和美式橄榄球)和不同比赛类型(NFL,世界杯和奥运会

等)的场地球类视频作为实验数据,以此来体现视频数据的多样性和广泛性。表 1 列举了实验中

所选择的视频数据,所有的视频都来自于国防科学技术大学人文与社会科学学院。

表 1 场地球类比赛视频
Tab.1 Field-ball game video

编号	名称	种类	比赛类型	格式
1	拜仁慕尼黑 vs. 斯伯丁	足球	欧洲冠军杯	MPEG
2	AC 米兰 vs. 锡耶纳	足球	意甲联赛	RMVB
3	阿森纳 vs. 赫尔城	足球	英超联赛	RMVB
4	巴西 vs. 法国	足球	世界杯	MPEG
5	中国 vs. 巴西	足球	奥运会	MPEG
6	南非 vs. 汤加	橄榄球	橄榄球世界杯	AVI
7	公羊 vs. 爱国者	美式足球	美国职业橄榄球联盟	AVI

数据的准备工作主要包括五个步骤:首先,将用于实验的视频数据分割成不同的视频片段,其中一半用于进行训练,另一半用于进行测试;第二,应用第 3 部分所提及的镜头探测和镜头分类算法对所有的训练视频片段进行处理,并根据类别平衡原则,从不同类型的镜头中提取相应数量的关键帧以构建训练数据集;第三,我们在测试视频片段中每隔 2s 提取三个连续的关键帧;第四,从所有的关键帧和视频片断的音频轨中提取数值特征;最后,通过人工来对每个关键帧进行标注,将有球门出现的“远景镜头”关键帧标记为“1”(Positive),其余的关键帧则标记为“0”(Negative)。

4.2 实验结果和分析

实验中所有的预处理代码都由 VC ++ 6.0, DirectX 9.0 SDK 和 OpenCV 来实现,其中 DirectX 已经被证明是多媒体处理中非常有效的解决方案。对于模型构建和分类工作,我们使用本实验室开发的 Salammbô 软件^[13]来设置模糊决策树算法的不同参数。

在经过预处理之后,我们得到含有 153 个属性(Attributes)的 960 个实例(Instances)用来训练,其中肯定的类别(Positive)和否定的类别(Negative)数量是相等的,同时也得到了具有相同属性的 960 个实例用来测试。所有的属性都是数值型数据。

为了便于实验结果的比较,我们同时实现了另外两种方法,即文献[3]中所提及的基于阈值的球门探测算法和 Weka3.6.0 中的 J48 决策树算法(Waikato Environment for Knowledge Analysis,它是一个由 Waikato 大学开发的开源程序,详见 <http://www.cs.waikato.ac.nz/~ml/index.html>)。在基于阈值的方法中,所有的训练实例都用来预

先确定阈值,然后基于已经确定好的阈值来对测试实例进行分类。而在机器学习方法(J48 决策树和模糊决策树)中,我们随机选择训练数据中的 2/3 用于训练模型,训练的模式设定为交叉验证(Cross-Validation),然后利用所获得的模型对测试实例进行分类。

表 2 是最终的实验结果。如果将表 2 第 3 行实验中得到基于模糊决策树的分类模型用于对单个类型的体育视频数据进行分类,则球门探测的查准率和查全率实验结果分别是:0.971 和 0.971(足球),0.943 和 0.938(橄榄球),0.935 和 0.934(美式橄榄球)。通过对实验结果的分析,得出以下结论:(1)相对于基于阈值的算法(F-Measure 的值是 0.86),监督式的机器学习方法(J48 的 F-Measure 值是 0.927,而模糊决策树的 F-Measure 是 0.962)能够较大地改进球门探测结果;(2)对训练数据进行平衡处理后,能够生成更加精确有效的分类模型(J48 中,非平衡训练数据的 F-Measure 值是 0.886,而平衡训练数据的 F-Measure 值是 0.927。在模糊决策树中,非平衡训练数据的 F-Measure 值是 0.893,平衡训练数据的 F-Measure 值是 0.962);(3)基于模糊决策树的算法还具有以下两个优点:

(1)当处理数值型数据时,相比于 J48,基于模糊决策树的球门探测算法得到了更加好的结果(J48 的 F-Measure 值是 0.927,而模糊决策树的 F-Measure 值是 0.962);

(2)还可以很容易地从 Salammbô 所构造的模糊决策树模型中推导出模糊规则,如,如果近绿色的颜色块的比率较大(大于 5 块),并且 MFCC 的平均值也很高(模糊集:大于 -0.7),那么当前帧包含球门;如果 Hough 直线的长度较长(模糊

表 2 场地球类比赛视频中球门探测的实验结果

Tab. 2 Comparative results of goalmouth detection in field-ball game video

算法	查准率		查全率		F-measure	
	平衡	非平衡	平衡	非平衡	平衡	非平衡
基于阈值的算法 ^[3]	0.883		0.836		0.86	
Weka 中的 J48 决策树算法	0.928	0.886	0.927	0.886	0.927	0.886
Salammbô 中的模糊决策树算法	0.962	0.894	0.962	0.893	0.962	0.893

集:大于 5 个像素),并且整个帧的颜色是近绿色,那么当前帧包含球门。

5 总结与展望

提出了一种基于模糊决策树的方法来探测场地球类比赛视频中的球门,实验结果表明,当与基于阈值的算法和决策树的算法相比时,提出的算法能够明显地改进分类的结果。同时,也能够很容易地从所建立的模糊决策树模型中推导出模糊规则,并以此来解释所生成的分类模型。

未来的研究工作将在以下几个方面开展:

- (1) 提取更加有效的底层特征来进行训练和分类;
- (2) 处理实时球门探测问题;
- (3) 将基于模糊决策树的球门探测算法集成到 DSVCAR 系统 (Digital Sports Video Content Analysis and Retrieval System, 软件著作权登记号 2009SR03486)。

参考文献 (References)

- [1] Yu X G, Farin D. Current and emerging topics in sports video processing [C]//Proceedings of International Conference on Multimedia and Expo. Amsterdam, Netherland; IEEE, 2005: 526 - 529.
- [2] Wan K W, Lim J, Xu X, et al. Real-time camera field view tracking in soccer video [C]//Proceedings of International Conference on Acoustics, Speech, and Signal Processing. Hong Kong, China; IEEE, 2003: 185 - 188.
- [3] Wan K W, Yan X, Yu X, et al. Robust goalmouth detection for virtual content insertion [C]//Proceedings of International Conference on Multimedia, Berkeley, USA; ACM, 2003: 468 - 469.
- [4] Marsala C, Detyniecki M. Fuzzy data mining for video [C]//Proceedings of European Society for Fuzzy Logic and Technology, Zittau, Germany, 2003: 73 - 78.
- [5] Fan X N, Xu L Z, Zhang X W, et al. The research and application of human detection based on support vector machine using in intelligent video surveillance system [C]//Proceedings of International Conference on Natural Computation, Jinan, China; IEEE, 2008: 139 - 143.
- [6] Zhao Z Q, Glotin H, Xie Z, et al. Cooperative sparse representation in two opposite directions for semi-supervised image annotation [J]. IEEE Transactions on Image Processing, 2012, 21(9): 4218 - 4231.
- [7] Zhao Z Q, Glotin H. Diversifying image retrieval by affinity propagation clustering on visual manifolds [J]. IEEE Multimedia, 2009, 16(4): 34 - 43.
- [8] Liang C H, Kuo J H, Chu W T, et al. Semantic units detection and summarization of baseball videos [C]//Proceedings of International Symposium on Circuits and Systems, Vancouver, Canada; IEEE, 2004: 297 - 300.
- [9] Zhang C Y, Pan B C, Zheng S L, et al. Motion object detection of video based on principal component analysis [C]//Proceedings of International Conference on Machine Learning and Cybernetics, Kunming, China; IEEE, 2008: 2938 - 2943.
- [10] Marsala C, Detyniecki M. UPMC-LIP6 at TrecVid '08: Balanced and unbalanced forests of fuzzy decision trees for high-level feature detection [R]. Gaitesburg; Proceedings of TREC Video Retrieval Evaluation, 2008.
- [11] Noor H, Mirza S H, Sheikh Y, et al. Model generation for video - based object recognition [C]//Proceedings of International Conference on Multimedia, Santa Barbara, USA; ACM, 2006: 715 - 718.
- [12] Yang Y, Lin S X, Zhang Y D, et al. Highlights extraction in soccer videos based on goal - mouth detection [C]//Proceedings of International Symposium on Signal Processing and Its Applications, Sharjah, UAE; IEEE, 2007: 1 - 4.
- [13] Marsala C, Bouchon - Meunier B. An adaptable system to construct fuzzy decision trees [C]//Proceedings of North American Fuzzy Information Processing Society Annual Conference, New York, USA; IEEE, 1999: 223 - 227.
- [14] Zadeh L A. Probability measures of fuzzy events [J]. Journal of Mathematical Analysis and Applications, 1968, 23(2): 421 - 427.
- [15] Bai L, Lao S Y, Liu H T, et al. Video shot boundary detection using Petri-Net [C]//Proceedings of International Conference on Machine Learning and Cybernetics. Kunming, China; IEEE, 2008: 3047 - 3051.
- [16] 卜江, 老松杨, 刘海涛, 等. 一种新的足球视频语义镜头表示及事件查询框架 [J]. 国防科技大学学报, 2009, 31(5): 50 - 53.
BU Jiang, LAO Songyang, LIU Haitao, et al. A new framework for semantic shot representation and event query of soccer video [J]. Journal of National University of Defense Technology, 2009, 31(5): 50 - 53. (in Chinese)