

体育视频分析

童晓峰 刘青山 卢汉清

(中国科学院自动化研究所模式识别国家重点实验室 北京 100190)

摘 要 体育视频因为拥有数量庞大的受众群体和巨大的商业应用前景而备受研究者和工业界的关注. 文中从底层特征提取、中级关键字生成、高级语义推理、相关应用研究和原型系统开发等方面, 综述了近年来体育视频分析的研究进展以及可能的发展趋势.

关键词 体育视频分析; 语义推理; 事件检测; 视频摘要; 特征提取

中图法分类号 TP391

A Survey on Sports Video Analysis

TONG Xiao-Feng LIU Qing-Shan LU Han-Qing

(National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190)

Abstract Sports video is a hot research topic for its wide viewer-ship and enormous application potential in recent years. This paper gives a review of relative research work including low-level feature extraction, mid-level keyword generation, high-level semantics inference, relative applications and system prototypes, and finally indicates potential trend.

Keywords sports video analysis; semantic inference; event detection; video summarization; feature extraction

1 引 言

体育视频是一类重要的媒体数据, 它拥有广大的观众群体和巨大的应用前景而受到学术和工业界的广泛关注. 随着移动设备和互联网的普及, 人们对体育视频也从直接观看和简单浏览转向多元化的需求, 如精彩片断摘要、特定事件检测、节目定制服务、视频内容编辑等. 这些服务都依赖于对体育视频进行语义分析与理解. 体育视频分析有着与一般视频处理相似的问题, 比如底层特征与高级语义之间存在的语义间隔; 也有自己的特别之处: (1) 体育领域中高级语义事件的定义比较明确, 减小了语义的主观性和模糊性; (2) 体育比赛有着特定的结构与规

则, 这些规则以及在视频广播中所采用的编辑方法都有助于视频的分析与理解.

体育视频分析的研究大约有十多年的历史, 目前取得了很大的进展, 也出现了一些应用模式和系统原型. 本文的目的就是综述近年来国内外体育视频分析的研究现状、遇到的问题和解决的办法, 希望帮助读者了解此方面的知识. 本文首先介绍体育视频分析的不同需求及研究内容; 然后讲述国内外当前研究现状, 包括框架、特征提取和算法分析等; 接下来描述了研究发展趋势; 最后给出了结论.

2 用户需求和研究内容

从对体育视频的需求、条件和可用资源等方面

来看,用户可以分为四类:电视用户、移动设备用户、网络用户和专业人士^[1].电视用户不担心数据传输带宽的问题.然而,他们可能不能及时观看比赛或者不能花几个小时观看比赛直播.为了节约时间并了解比赛情况,他们对体育视频摘要比较感兴趣.随着3G无线标准的产生和应用,移动用户有快速的网络连接.但是由于带宽的限制,实时的数据流传输仍然不现实,另外体育节目的价值随着时间推移而降低,他们希望及时了解比赛进程.所以,移动用户关心实时的精彩片断提取与传送.网络用户与移动用户一样关心网络带宽问题,他们的需求包括视频摘要和特定感兴趣事件检测.专业用户包括运动员、教练员和体育评论者.他们需要准确提取球队和运动员的某些信息以制定比赛计划、评估队员的表现或者分析比赛策略.这些用户对目标检测与跟踪、运动轨迹提取以及在此基础之上的语义分析感兴趣.

体育视频分析在不同需求的推动下产生了很多有价值的应用,包括精彩片断提取与传输、视频摘要、视频浏览与检索、球和运动员的检测与跟踪、行为与动作分析及索引、战术统计与策略分析、虚拟内容插入以及虚拟场景构造等等.

3 研究现状

3.1 框 架

体育视频分析存在底层特征与高级语义之间的语义间隔问题.目前的办法是构建一个中间描述层作为低级和高级语义之间的桥梁,在构建中级描述

层的时候,加入先验知识和特定领域相关规则,辅助底层特征选择和高级语义推理.通常采用的框架是一个三层次结构^[2],即低级特征层、中级语义描述层和高级事件层,如图1所示.低级特征层包括基本的视觉(比如颜色、形状、纹理、运动等)、听觉(比如LPC、MFCC、LPCC、STE等)和文本特征,它们可以直接从视频数据中提取.框架的高层是一些语义实体,比如比赛的结构和内容、精彩片断、特定语义事件等.“事件”被定义为用户感兴趣的具有一定上下文线索并符合特定领域知识模型的语义时空实体^[3].中间层定义了对视频片段的描述,称为关键字,包括:(1)视频关键字,比如镜头类型^[4-5]、运动模式^[6-9]、纹理与形状描述^[10]、比赛位置^[11-12]、目标位置与轨迹等;(2)音频关键字,比如哨声、解说员声音、观众欢呼声以及静音等^[13-14]; (3)文本关键字,比如“进球”、“犯规”等,文本包括场景和人工叠加的字幕(caption)^[15]、声音转录字幕(close-caption)^[16]以及网上广播文字^[17-18]等.

体育节目的中级关键字提取和高级语义推理一般需要结合特定领域知识以及视频编辑规则.原因在于:(1)很难自动找出适合于某些高级语义事件推理的底层特征;(2)底层特征和高级语义存在距离,底层特征难以直接描述高级语义;(3)高级语义通常是一个时空实体,拥有时间上和空间上的跨度.底层特征表达式维度高而复杂,而且难以表达语义的不确定性.领域特定知识可以引导底层特征选择和提取,生成中级关键字,再结合不同语义事件的特点选择推理算法.

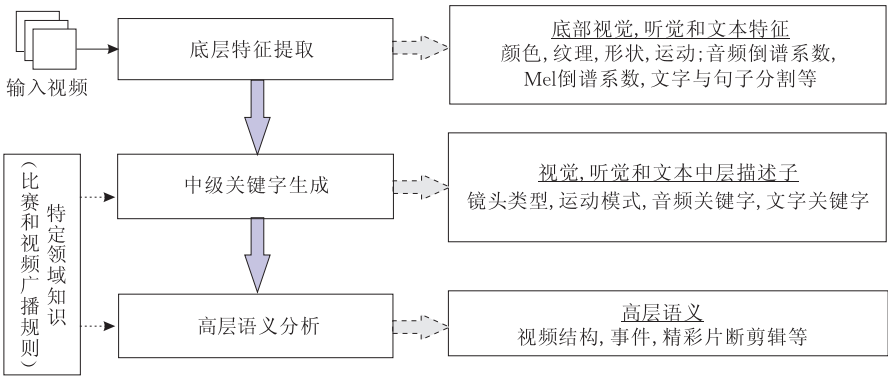


图1 体育视频分析框架

通常讲的视频包含视觉、听觉和文本等三个方面的信息源.下面,我们首先从这三个方面回顾一下体育视频分析中底层特征提取和中级关键字生成,然后介绍高级语义分析,研究方法和发展路线.

3.2 底层特征的提取

3.2.1 视觉特征

视觉信息是体育视频分析中一个重要成分.早期的很多工作都是利用视觉信息来完成的^[19-20].视

觉特征主要有颜色、纹理、形状和运动等。基于颜色的处理有视频主颜色提取^[1]、镜头分类以及特定目标表征、检测、跟踪与识别^[21-22]等。纹理特征通常用于画面的分类^[5]、特定目标描述与检测^[23]等。形状特征可用于目标表达、比赛场地描述和运动姿态识别等等。比如,描述球的形状^[24-25]、场地形状与位置判断^[12,26]等。运动特征对于表征体育视频非常重要,运动模式反映了比赛的节奏。文献[9]用运动活度(motion activity)表达比赛的节奏,从而提取比赛激烈的片段。Huang 等^[27]利用运动特征来描述重放场景的 logo 过程以实现重放场景的自动学习与检测,取得了良好的效果。

3.2.2 听觉特征

音频蕴涵了丰富的语义,近年来很多方法开始采用音频特征来分析和检索多媒体数据。相对视频信号,音频处理速度快,计算量较小。所以,音频处理可以用于快速的事件时刻定位,然后结合视觉特征进一步处理。从音频例子中提取特征,最简单的方式是直接从音频例子(audio clip)中提取时域和频域特征来表征音频例子所蕴涵的语义。常用的音频特征有 mel 频率倒谱系数(MFCC)、短时能量(STE)、过零率(ZCR)和线性预测系数(LPC)等特征,最后把它们的统计量(如均值和方差等)计算出来作为音频例子的特征向量。Li 等^[12]利用音频 MFCC 特征和混合高斯模型来检测兴奋语音和哨音。Xu 等^[28]综合利用了 MFCC、STE 等特征和 SVM 对长短哨音、兴奋语音等进行识别。

3.2.3 文字特征

文字也是多媒体数据中的一个重要信息源,传统的文字数据有画面字幕和转录字幕。画面字幕包括场景字幕和人工字幕。场景字幕由于检测和识别比较困难,难以应用,但是人工字幕易于提取和识别并用于视频分析和理解。文献[15]利用字幕检测比赛记分牌上的信息以实现棒球节目中的事件检测。转录字幕是通过语音识别技术转录生成的文字,如新闻解说和场景对话等。可以采用传统文本分析技术识别其中的关键字,实现多媒体数据分析与理解^[18,29]。另一种新的文字信息——网上直播文字,在网络上用文字及时描述比赛实况,它包含发生的

事件、事件的主体、动作及结果等。直播文字可以用于实时检测比赛精彩片段^[17-18,30]等工作。

3.3 中级关键字提取

很多高级语义事件蕴涵了特定的语义场景,这些场景往往与特定领域知识相关。中级关键字作为桥梁连接了底层特征和高级语义,填补了它们之间的语义间隔。体育视频中常用的中级关键字包括镜头类型、重放场景、比赛位置、特定目标位置与轨迹、运动模式、音频关键字与文本关键字等。

3.3.1 语义镜头分类

比赛语义场景通过镜头类型及其转换上下文来表达。相应地,特定镜头类型及其转换上下文预示着特定语义事件^[5,31-32]。镜头类型包括长镜头、中镜头、特写和场外镜头等。长镜头一般用于显示比赛场地的全貌,中镜头可以表现运动员的动作,而短镜头(特写)则可以近距离地刻画人物的表情,场外画面可以描述场外观众的行为与反应。

Ekin^[1]根据场地面积比和场地分布等特征实现对足球和篮球视频的镜头分类。Xu 等^[33]利用场地面积和场地中的物体尺寸等信息对足球视频进行镜头分类,并把不同镜头类型映射到不同比赛状态以对足球视频进行结构分析,将其分为比赛和中断(play/break)两个状态。Duan 等^[2,5]根据场地面积、场地线和场地中的物体尺度等特征结合相应的领域知识利用决策树实现对足球、篮球、网球、排球等节目的镜头分类。文献[31]统计了颜色矩与形状信息并使用 HMM 模型对台球和网球节目镜头进行分类。

3.3.2 重放场景检测

在广播视频中,精彩的片断一般会用慢速播放方式从不同视角重复播放几次。因此,重放场景对于精彩片断检测有非常重要的提示作用。重放场景分为两类:不带 logo 的重放场景和带 logo 的重放场景。后者是指在重放场景的开始和结束处有一个实况场景与重放场景之间的转换,这个转换过程通常会出现一个 logo,比如足球世界杯的徽标、奥运会徽标等,如图 2。大部分工作都是检测有 logo 的重放场景^[27,34-37]。在对 logo 的特征表达上,文献[34-35]使用了颜色特征,但是实验表明^[27]采用运动特征得



图 2 带有 logo 的重放场景转换

到的结果更好,运动特征更能描述 logo 转换的本质.对于不带 logo 的重放场景,特征提取和模式描述都比较困难,相应的工作较少^[26,38].

3.3.3 比赛位置判断

比赛位置是指场地球类运动中当前画面对应的场地上的位置,比如左边禁区前沿、中场等.比赛位置及其位置转换表达了比赛的节奏和状态.另外精彩镜头比如进球或射门时比赛位置一般在禁区前沿,因此它也作为一个中级关键字检测语义事件^[12,26].

判断比赛位置需要检测场地中的直线并且识别这些直线或者通过统计学习方法推导场地位置.简单的操作可以把场地分为左半场、中场还是右半场^[23].再复杂一些可以分为 5 个区域^[12],如图 3.更加细致的是 15 个区域^[26],但是运算更复杂,准确性也不高.实验证明,5 个区域已经足够检测像射门、进球、角球等事件^[12-13].

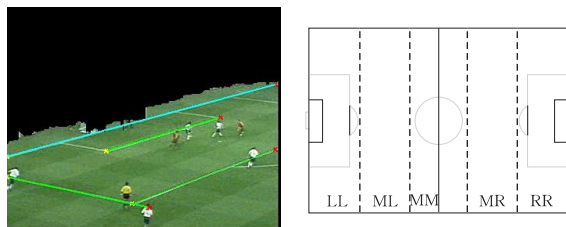


图 3 Hough 直线检测以及比赛位置划分

3.3.4 目标检测和跟踪

在体育节目中,另一个令人关注的任务是目标检测、跟踪与轨迹分析.这些目标包括运动员、裁判、球和球门等.检测这些语义目标的目的,一方面是为了分析高级语义事件,得到视频摘要;另一方面是为了比赛体育策略分析、行为分析与索引等,比如比赛阵形、进攻路线以及配合动作等.

3.3.4.1 球的检测与跟踪

早期的足球检测基本上都采用基于颜色的模板匹配技术^[19-20],跟踪则使用模板匹配或者 Kalman 滤波方式^[25].由于足球的尺寸较小,速度较快,遮挡较多,基于单帧或前后几帧图像的算法难以实现稳定有效的检测与跟踪.Yu 等^[22]提出了一种新的思路:在检测阶段并不确定一个目标,而是得到多个候选点,然后对多个候选点的运动轨迹进行推理,确定最优轨迹.文献^[39-40]也采取了类似的思路,先检测跟踪多个候选目标,然后通过轨迹优化得到最佳目标及其运动路径.

3.3.4.2 运动员的检测、跟踪与分类

运动员的检测、跟踪与分类是策略分析中重要

模块,但是由于运动员众多、数目变化、遮挡拥挤、姿态变化、表观特征区分性不强等因素,这个任务也具有挑战性.运动员跟踪主要是多个目标协同跟踪,分类是对将场上运动员(包括裁判)进行身份判别:是哪一个球队或是裁判员.Wang 等^[41]初步试探了两方球员球衣颜色的自动建模与检测.文献^[21]等用场地分割图像作为掩模采用 AdaBoost 方法检测,性能较之前的方法提高很大.还通过自动的运动员检测搜集样本,利用无监督聚类学习球衣颜色,然后对运动员进行分类.这个方法实现了对双方球员和裁判的自动高性能分类.

3.3.4.3 球门的检测

球门是一个特殊的目标,它对于射门和进球等事件有提示作用.检测球门的目的主要有两个方面的应用:(1)作为一个线索检测射门事件或精彩镜头;(2)作为有意义的目标进行视频内容编辑,比如虚拟广告叠加.Wan 等^[42]根据球门的底部总是在场地边线上的先验知识,先检测场地边线,然后通过灰度增长搜索检测门柱和横梁.当以上 3 个要素都检测到后,就认为出现球门.这个方法速度较快,效果也很好.

3.3.5 运动模式

运动信息表示了视频图像内容在时间轴上的发展变化,它可用于视频分类、检索和语义内容理解等.不同类型的视频会有不同的运动模式,利用运动信息可以对视频进行分类.比如,利用主元分析法(PCA)对视频的运动向量进行分析,可以将不同类的体育比赛的片断进行分类^[43].通过检测视频中重放慢镜头和摄像机的运动可对体育运动视频和非体育运动视频的混合数据库进行分类^[9].

在体育视频处理中,一般将运动信息和其它特征结合起来进行分析.文献^[2]将运动信息作为中级描述子之一对视频的高级语义理解进行推理和分析.文献^[44]中利用运动活跃度(motion activity)、镜头密度和语音能量等 3 个特征来检测精彩片断.文献^[45]利用全局运动信息来检测和检索不同球队的攻防场景.

3.3.6 音频关键字

音频信息可实现对特定语义的场景的检测,比如任意球/点球、犯规、得分、射门和比赛开始/结束等^[28,46-48].相对于视频信号而言,它计算量小,处理速度快.而且有些事件用音频检测更有效,比如对于犯规事件而言,哨音是一个重要有效特征.在体育视频中,音频中级关键字包括哨音(长哨音、双哨音和

多哨音),解说员语音(兴奋和平静的)以及观众声音(兴奋和平静的).Rui 等^[49]仅仅利用音频特征实现了棒球节目的精彩片断检测.文献[12]等融合了音频和视频信号检测精彩片断.文献[50]采用了隐马尔可夫模型(HMM)对上述音频关键字进行了分类识别,结果显示比 SVM 分类器有更好的效果.文献[14,51]将音频特征和视觉特征融合起来进行快速的精彩片断提取.

3.3.7 其它中级关键字

除了视觉和听觉信息之外,另一个重要的信息源是文字信息.在体育视频中,文字信息也被用来进行基于语义事件检测与分析^[15-16,52-53].从文本信息中识别、寻找或者挖掘一些关键字,比如“进球”、“头球”、“越位”等等,然后结合视频或音频信息定位并得到相应事件发生的视频段.

3.4 高级语义分析

体育视频处理与分析的应用主要表现在 3 个方面:(1)对体育节目进行自动的结构分析,便于存储、管理和检索;(2)语义事件分析,分析理解体育节目中有意义的事件,便于语义摘要、浏览和索引;(3)经过分析对其中的内容进行编辑和丰富.

3.4.1 视频结构分析

视频结构分析可以实现视频的语义结构化解析.在视频中,镜头是对视频流进行处理的基本单元.在视频处理中,首要的步骤是找到镜头的切分点.镜头边缘检测的方法有绝对帧间差法、像素差法、数值差法、颜色直方图法、边缘差法、压缩差法和运动矢量法等.实际上在体育视频中,除了重放场景转换中的镜头渐变外,其它大多数镜头切换都是突变.因此,渐变检测一般用于重放场景检测,而对于突变检测,一般的基于颜色直方图阈值分割的算法就可以达到比较好的效果.

对于体育视频而言,一种特殊的结构分类方式,就是比赛进行/中断(Play/break).这种结构化分析有两个好处:(1)可以剔除非比赛状态的段落,只保留比赛进行时的视频,从而节约存储空间;(2)这两种状态以及之间的转换蕴涵一定的语义.Xu 等^[33]在镜头分类的基础上分析 Play/break 结构.Xie 等^[54]提取了场地面积比和运动活度两个特征,并使用隐马尔可夫模型和动态规划优化方法提高了 Play/break 结构分析的准确度.Duan 等^[2]在画面分类的基础上,检测了体育节目的 Play/break 结构,并将这种结构分类推广到几种场地球类运动中,比如足球、篮球、网球、排球等.

不同体育节目有不同的结构,结合领域知识可以得到更具体、有意义的结构.比如在网球节目中,其结构有分(point)、局(game)、盘(match)等语义结构.Zhang 等^[55]提出了网球比赛的多层次语义结构模型.它结合了领域知识和监督学习方法检测了网球中的发球场景和棒球中的击球场景.文献[23,56]提出足球比赛单元(play-unit)的概念,实际上它相当于比赛中的一个回合,从发球开始到该球死掉.

3.4.2 事件检测

大多数用户关注的是精彩事件摘要和特定感兴趣事件浏览,随着需求扩大和研究深入,研究内容扩展到视频摘要、视频片段(镜头或事件)索引、特定事件检测(如进球、角球、任意球、犯规等)、目标检测与跟踪(比如球、运动员等)以及行为与策略分析等等.

精彩事件一般指与得分、进球等相关的场景、动作及其上下文,但是很难给出一个统一准确的定义,比如足球、篮球中一般指射门得分,网球中的 aces 球以及精彩击球,橄榄球中的触地得分等.精彩片断在广播节目中的表现手法基本相同,一般伴随着观众的欢呼和慢运动重放等场景.Rui^[49]和 Radhakrishnan^[57]利用音频特征检测棒球中精彩的精彩场景,Assfalg 在镜头分类的基础上利用 HMM 实现了足球节目精彩片断提取^[58],Wan 等开发了针对移动用户的精彩片断提取方法^[59]等等.

精彩片断可用于视频摘要,其应用有:(1)将精彩片断按照时间的先后顺序串联起来作为视频摘要^[60].这种方式的好处是可以根据摘要的长度来定制不同的视频摘要.(2)提取出精彩片断,对它们的精彩程度进行排序,然后按照精彩程度高低组织起来^[61],它可以让用户优先浏览到精彩程度较高的片断.

特定语义事件检测也是一个活跃的方向.比如检测足球中的射门、进球、角球、任意球、越位、红黄牌事件等.特定事件都有特定的模式和语义上下文.虽然在特定的事件中出现的场景不一定完全相同,但是基本上模式(场景和场景之间的转换)比较接近.因此,目前基于场景上下文到语义内容的思想(from context to content)并利用概率统计模型来检测特定语义事件的思路得到广泛认可.特定语义事件检测可应用于:(1)在视频中自动定位感兴趣事件片断;(2)对视频进行以事件类型为索引的结构化;(3)为用户提供按照特定事件快速智能浏览.

在体育视频分析中另一个比较重要的方向是特定目标检测与跟踪以及语义和策略分析,比如球以

及运动员等的检测与跟踪。得到球的位置后,可以在拼接的全景图中表示进球的路线与过程^[20]。球与球门之间的距离还可以用于判断可能的射门事件的发生^[22]。通过运动员检测、跟踪和分类以及场地配准,将运动员的轨迹映射到球场模型上,可以分析球队阵形、进攻路线、传球路线等行为与策略等^[11,21,62-63]。

3.4.3 内容编辑与增强

在视频分析和理解的基础上,可以对其内容编辑和增强,比如虚拟内容插入、三维场景构造以及自动视频广播与编辑。

虚拟内容插入^[42,64-65]是指在原视频画面中插入虚拟内容,这些内容包括图像、视频和文字等信息。虚拟内容插入具有代价小、成本低、灵活方便、不受实际场地和时间限制等优点。技术关键是要选择在合适的时间和合适的地点插入合适的内容,使得插入的内容既达到预期目的,又不影响观众的正常观看。

基于视频的三维场景构造^[66-69]先虚拟构造比赛场景,然后传送运动或感兴趣的内容,比如足球比赛中的运动员、球等目标。这项工作需要在实际场景中判断比赛发生的地点,分割出比赛场地和运动目标,然后传送特定内容,或者通过变换将实际场景变换到特定视角来播放。它既可以保留节目中感兴趣部分,又可以节约传输数据、时间和成本。

目前的广播视频编辑都是通过专业人员人工完成的,有着很强的主观依赖性。半自动/自动编辑是一个需求。目前自动节目编辑方面的工作比较少,已有的工作包括在主摄像机环境下的自动重放场景插入^[70]、自动的体育音乐视频编辑^[71]等。

4 发展趋势

体育视频分析经过十多年的发展,在研究内容、算法、系统和应用几个方面都有很大扩展与进步。从研究的广度来说,从单一的足球节目到多项球类运动(足球、篮球、网球、排球、乒乓球、橄榄球、棒球等)、田径运动以及游泳等运动。在算法方面,由单一模态特征到多模态特征融合,从启发式推理到自动机器学习,算法在性能和速度上都有很大进步。

4.1 多模态信息融合

基于单模态特征的分析方法只利用视觉、听觉或者文本信息三者之一的信息来处理和分析视频,没有充分有效使用有效信息。多媒体信息流本质上是由文本、图像、图形、音频和视频等多态媒质交互

融合形成的,综合利用多模态特征才能完整表示多媒体所蕴涵的语义信息。多模态信息融合一方面充分利用了媒体的各个方面的数据,实现不同的分析和应用;另一方面在多个模态之间互补,提高语义分析精度,加快处理速度或者检测更多语义。在多模态信息融合方面需要考虑3个问题:(1)融合哪些模态信息。不同的应用场景需要不同的模态信息,要合理选择;(2)在什么层次上融合。可以在底层特征层次、中级关键字层和高级语义级别上的融合;(3)如何融合,包括数据规范化处理、融合概率表达等。目前多模态特征融合的工作有视觉信息和听觉信息的融合^[2,14,51,59,72-74]、视觉信息与文本的结合^[9,15,29,52,75]以及视频、音频和文字三者的结合^[18,71,76]。

4.2 机器学习的应用

在早期的工作中,利用领域相关先验知识和启发式推理可以方便快捷的得到部分事件检测结果^[2,28,33,77]。这种方式简单易用,不过主观性较强,推广能力较差。很多语义事件并非确定性的,利用启发式推理难以实现自动多种事件检测。目前模式识别和机器学习技术被广泛应用于视频分析的各个阶段,比如底层特征选择、中级关键字提取与融合、底层特征或中级关键字到高级语义的推理。涉及到的方法有贝叶斯网络^[78-80]、动态贝叶斯网络^[78]、支撑向量机^[14,28]、隐马尔可夫模型^[36,50,81-82]、最大熵模型^[73]、决策树^[5]、神经网络和有限状态机^[83]等等。

4.3 处理的实时性

体育节目有很强的时间性,如果不能实时处理并得到分析结果,就有可能失去应用价值。目前很多方法没有注意这个方面,只是侧重于处理的精确度以及能分析的内容。一个实用的体育视频分析系统应该满足有效和高效两个能力。实时性一方面要求我们采用更有效的特征和更高效的算法,另一方面随着多核时代的到来,并行计算也是大势所趋。

4.4 与通信技术相结合

作为实用的应用研究原型,体育视频分析的结果要方便地提供给各种接入和获取方式的用户使用。对于电视用户来说,只需要简单地把处理结果播放出来就可以了。但是,对于移动用户和网络而言,由于带宽以及费用的限制,他们需要准确有价值又收费合理的信息提供方式,所以不同应用场景下的视频编码、传输以及分发方案都需要考虑,这些要求需要与相关通信技术结合起来。

5 结束语

体育视频分析是一个年轻而充满活力的研究领域. 目前关于体育视频处理、分析和编辑各个方面的研究工作都已经展开. 体育视频既有传统视频语义分析的问题和困难, 同时作为一项拥有大量观众和巨大市场前景的特殊媒体, 它有着特别的需求和处理方法. 虽然目前国内外众多研究者在不同方向做了大量工作, 出现了一些原型系统, 不过, 它离实际应用还有一定距离. 一方面, 我们还需要研究更加高级和实用的技术, 另一方面, 需要结合实际应用场景和需求开发出实用的系统, 使这些技术能够真正为人类服务.

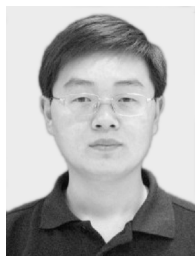
参 考 文 献

- [1] Ekin A, Tekalp A, Mehrotra R. Automatic soccer video analysis and summarization. *IEEE Transactions on Image Processing*, 2003, 12(7): 796-807
- [2] Duan L, Xu M, Chua T, Tian Q, Xu C. A mid-level representation framework for semantic sports video analysis//*Proceedings of the ACM Multimedia*. Berkeley, CA, USA, 2003: 33-44
- [3] Sun Xing-Hua, Xu Guang-You. Research on video retrieval based on semantic events. Post-doctoral Research Report, Tsinghua University, Beijing, 2003(in Chinese)
(孙兴华, 徐广佑. 基于语义事件的视频检索研究. 清华大学博士后研究报告, 北京, 2003)
- [4] Duan L, Xu M, Yu X, Tian Q. A unified framework for semantic shot classification in sports videos//*Proceedings of the ACM Multimedia*. Juan-les-Pins, France, 2002: 419-210
- [5] Duan L, Xu M, Tian Q. Semantic shot classification in sports video//*Proceedings of the SPIE Storage and Retrieval for Media Databases*. San Jose, USA, 2003, 5021: 300-313
- [6] Cheng F, Christmas W, Kittler J. Periodic human motion description for sports video databases//*Proceedings of the International Conference on Pattern Recognition*. Cambridge, UK, 2004: 870-873
- [7] Cutler R, Davis L. Robust real-time period motion detection, analysis, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 781-796
- [8] Duan L, Xu M, Tian Q, Xu C. Nonparametric motion model with application to camera motion pattern classification//*Proceedings of the ACM Multimedia*. New York, USA, 2004: 328-331
- [9] Kobla V, DeMenthon D, Doermann D. Identification of sports videos using replay, text, and camera motion features// *Proceedings of the SPIE Storage and Retrieval for Media Database*. San Jose, USA, 2000, 3972: 332-343
- [10] Ma Y, Zhang H. Motion texture: A new motion based video representation//*Proceedings of the International Conference on Pattern Recognition*. Barcelona, Spain, 2000: 548-551
- [11] Iwase S, Saito H. Parallel tracking of all soccer players by integrating detected position in multiple view images//*Proceedings of the International Conference on Pattern Recognition*. Cambridge, UK, 2004: 751-754
- [12] Li J, Wang T, Hu W et al. Two-dependence Bayesian network for soccer highlight detection//*Proceedings of the IEEE International Conference on Multimedia and Expo*. Toronto, Canada, 2006: 1625-1628
- [13] Wang T, Li J et al. Semantic event detection using conditional random fields//*Proceedings of the Workshop on Semantic Learning Applications in Multimedia*. New York, USA, 2006: 611-618
- [14] Xu M, Duan L, Xu C, Tian Q. A fusion scheme of visual and auditory modalities for event detection in sports video// *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Hong Kong, China, 2003: 105-115
- [15] Zhang D, Chang S-F. Event detection in baseball video using superimposed caption recognition//*Proceedings of the ACM Multimedia*. Juan-les-Pins, France, 2002: 315-318
- [16] Nitta N, Babaguchi N. Automatic story segmentation of closed-caption text for semantic content analysis of broadcasted sports video//*Proceedings of the International Workshop on Multimedia Information Systems*. Arizona, USA, 2002: 110-116
- [17] Dai J, Duan L, Tong X, Xu C, Tian Q, Lu H, Jin J. Replay scene classification in soccer video using web broadcast text// *Proceedings of the IEEE International Conference on Multimedia and Expo*. Amsterdam, Netherlands, 2005: 1098-1101
- [18] Xu H, Chua T-S. The fusion of audio-visual features and external knowledge for event detection in team sports video// *Proceedings of the ACM Workshop on Multimedia Information Retrieval*. New York, USA, 2004: 127-134
- [19] Gong Y, Sin L, Chuan C, Zhang H, Sakauchi M. Automatic parsing of TV soccer programs//*Proceedings of the International Conference on Multimedia Computing and Systems*. Washington, DC, USA, 1995: 167-174
- [20] Yow D, Yeo B, Yeung M, Liu B. Analysis and presentation of soccer highlights from digital video//*Proceedings of the Asian Conference on Computer Vision*. Singapore, 1995: 499-503
- [21] Liu J, Tong X et al. Automatic player detection, labeling and tracking in broadcast soccer video//*Proceedings of the British Machine Vision Conference*. UK, 2007: 70-80
- [22] Yu X, Xu C et al. Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video//*Proceedings of the ACM Multimedia*. Berkeley, CA, USA, 2003: 11-20

- [23] Tong X, Liu Q et al. A unified framework for semantic shot representation of sports video//Proceedings of the ACM Multimedia Information Retrieval. Singapore, 2005: 127-134
- [24] Orazio T, Ancona N, Canada G, Nitti M. A ball detection algorithm for real soccer image sequence//Proceedings of the International Conference on Pattern Recognition. Quebec, Canada, 2002: 210-213
- [25] Seo Y, Choi S, Kim H, Hong K. Where are the ball and players? Soccer game analysis with color-based tracking and image mosaic//Proceedings of the International Conference on Image Analysis and Processing. Florence, Italy, 1997: 196-203
- [26] Wang J, Chng E, Xu C. Soccer replay detection using scene transition structure analysis//Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Montreal, Canada, 2004: 433-437
- [27] Huang Q, Hu J, Hu W et al. A reliable logo and replay detector for sports video//Proceedings of the IEEE International Conference on Multimedia and Expo. Beijing, China, 2007: 1695-1698
- [28] Xu M, Maddage N et al. Creating audio keywords for event detection in soccer video//Proceedings of the IEEE International Conference on Multimedia and Expo. Baltimore, USA, 2003: 281-284
- [29] Babaguchi N, Kawai Y, Kitahashi T. Event based indexing of broadcasted sports video by intermodal collaboration. IEEE Transactions on Multimedia, 2002, 4(1): 68-75
- [30] Xu C, Wang J et al. Live sports event detection based on broadcast video and web-casting text//Proceedings of the ACM Multimedia. Santa Barbara, USA, 2006: 221-230
- [31] Dahyot R, Rea N, Kokaran A. Sports video shot segmentation and classification//Proceedings of the SPIE Visual Communications and Image Processing. Lugano, Switzerland, 2003, 5150: 404-413
- [32] Ide I, Yamamoto K, Tanaka H. Automatic video indexing based on shot classification//Proceedings of the International Conference on Advanced Multimedia Content Processing. Osaka, Japan, 1999: 87-102
- [33] Xu P, Xie L et al. Algorithms and system for segmentation and structure analysis in soccer video//Proceedings of the IEEE International Conference on Multimedia and Expo. Tokyo, Japan, 2001: 721-724
- [34] Tong X, Liu Q, Lu H. Replay detection in broadcasting sports videos//Proceedings of the International Conference on Image and Graphics. Hong Kong, China, 2004: 337-340
- [35] Duan L, Xu M, Tian Q, Xu C. Mean shift based video segment representation and applications to replay detection//Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Montreal, Canada, 2004: 709-712
- [36] Pan H, Beek P, Sezan M. Detection of slow-motion replay segments in sports video for highlights generation//Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Salt Lake City, USA, 2001: 1649-1652
- [37] Pan H, Li B, Sezan M. Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions//Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Orlando, USA, 2002: 3385-3388
- [38] Kobla V, Doermann D. Detection of slow-motion replays for identify sports videos//Proceedings of the IEEE Workshop Multimedia Signal Processing. Copenhagen, Denmark, 1999: 135-140
- [39] Tong X, Wang T, Li W, Zhang W. A three-level scheme for real-time ball tracking//Proceedings of the Workshop Multimedia Content Analysis and Mining. Weihai, China, 2007: 161-171
- [40] Yan F, Kostin A, Christmas W, Kittler J. A novel data association algorithm for object tracking in clutter with application to tennis video analysis//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA, 2006: 634-641
- [41] Wang L, Zeng B et al. Automatic extraction of semantic colors in sports video//Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Montreal, Canada, 2004: 617-620
- [42] Wan K, Yan X, Yu X, Xu C. Real-time goal-mouth detection in MPEG soccer video//Proceedings of the ACM Multimedia. Berkeley, CA, USA, 2003: 311-314
- [43] Sahouria E, Zakhori A. Content analysis of video using principal components. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(12): 1290-1298
- [44] Hanjalic A. Generic approach to highlights extraction from a sport video//Proceedings of the IEEE Conference on Image Processing. Barcelona, Spain, 2003: 14-17
- [45] Jiang S, Liu H, Zhao Z et al. Generating video sequence from photo image for mobile screens by content analysis//Proceedings of the IEEE International Conference on Multimedia and Expo. Beijing, China, 2007: 1475-1478
- [46] Liu Z, Wang Y, Chen T. Audio feature extraction and analysis for scene segmentation and classification. Journal of VLSI Signal Systems for Processing for Signal, Image, and Video Technology, 1998, 20(1): 61-80
- [47] Zhang T, Kuo C. Heuristic approach for generic audio data segmentation and annotation//Proceedings of the ACM Multimedia. Orlando, USA, 1999: 67-76
- [48] Wu C, Ma Y, Zhang H, Zhong Y. Event recognition by semantic inference for sports video//Proceedings of the IEEE International Conference on Multimedia and Expo. Lausanne, Switzerland, 2002: 805-808
- [49] Rui Y, Gupta A, Acero A. Automatically extracting highlights for TV baseball programs//Proceedings of the ACM Multimedia. Los Angeles, USA, 2000: 105-115

- [50] Xu M, Duan L et al. HMM-based audio keyword generation//Proceedings of the Pacific-Rim Conference on Multimedia. Tokyo, Japan, 2004: 566-573
- [51] Wan K, Xu C. Efficient multimodal features for automatic soccer highlight generation//Proceedings of the International Conference on Pattern Recognition. Cambridge, UK, 2004: 973-976
- [52] Babaguchi N, Nitta N. Intermodal collaboration: A strategy for semantic content analysis for broadcasted sports video//Proceedings of the IEEE International Conference on Image Processing. Barcelona, Spain, 2003: 13-16
- [53] Zhang D, Raj R, Chang S-F. General and domain-specific techniques for detecting and recognizing superimposed text in video//Proceedings of the IEEE International Conference on Image Processing. New York, USA, 2002: 593-596
- [54] Xie L, Chang S-F, Divakaran A, Sun H. Structure analysis of soccer video with hidden Markov models//Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Orlando, USA, 2002: 4096-4099
- [55] Zhang D, Chang S-F. Structure analysis of sports video using domain models//Proceedings of the IEEE International Conference on Multimedia and Expo. Tokyo, Japan, 2001: 713-716
- [56] Wang L, Le M, Xu G. Offense based temporal segmentation for event detection in soccer video//Proceedings of the ACM Workshop on Multimedia Information Retrieval. New York, USA, 2004: 259-266
- [57] Radhakrishnan R, Xiong Z, Divakaran A, Ishikawa Y. Generation of sports highlights using a combination of supervised and unsupervised learning in audio domain//Proceedings of the International Conference on Information, Communications and Signal Processing. Singapore, 2003: 953-959
- [58] Assfalg J, Bertini M, Bimbo A, Nunziati W, Pala P. Soccer highlights detection and recognition using HMMs//Proceedings of the IEEE International Conference on Multimedia and Expo. Lausanne, Switzerland, 2002: 825-828
- [59] Wan K, Wang J, Xu C, Tian Q. Automatic sports highlights extraction with content augmentation//Proceedings of the Pacific-Rim Conference on Multimedia. Tokyo, Japan, 2004: 19-26
- [60] Li B, Errico J, Pan H, Sezan M. Bridging the semantic gap in sports//Proceedings of the SPIE Storage and Retrieval for Media Database. San Jose, USA, 2003, 5021: 314-326
- [61] Tong X, Liu Q, Zhang Y, Lu H. Highlight ranking in sports video//Proceedings of the ACM Multimedia. Singapore, 2005: 519-522
- [62] Pascual F, Neucimar L et al. Tracking soccer player using the graph representation//Proceedings of the International Conference on Pattern Recognition. Cambridge, UK, 2004: 787-790
- [63] Iwase S, Saito H. Tracking soccer players based on homography among multiple views//Proceedings of the SPIE Visual Communications and Image Processing. Lugano, Switzerland, 2003, 5150: 283-292
- [64] Xu C, Wan K, Bui S, Tian Q. Implanting virtual advertisement into broadcast soccer video//Proceedings of the Pacific-Rim Conference on Multimedia. Tokyo, Japan, 2004: 264-271
- [65] Das S, Rosser R, Tan Y. Method of tracking scene motion for live video insertion systems. US Patent 5808695, 1998
- [66] Liang D, Liu Y et al. Video2Cartoon: Generating 3D cartoon from broadcast soccer video//Proceedings of the ACM Multimedia. Singapore, 2005: 217-218
- [67] Yan X, Yu X, Hay T. A 3D reconstruction and enrichment system for broadcast soccer video//Proceedings of the ACM Multimedia. New York, USA, 2004: 746-746
- [68] Bebie T, Bieri H. A video-based 3D-reconstruction of soccer games. Eurographics, 2000, 19(3): 391-400
- [69] Koyama T, Kitahara I, Ohta Y. Live 3D video in soccer stadium//Proceedings of the ACM SIGGRAPH 2003 Sketches and Applications. San Diego, USA, 2003: 178-187
- [70] Wang J, Xu C et al. Automatic replay generation for soccer video broadcasting//Proceedings of the ACM Multimedia. 2004: 11-20
- [71] Wang J, Xu C et al. Automatic generation of personalized music sports video//Proceedings of the ACM Multimedia. Singapore, 2005: 735-744
- [72] Chang Y, Zeng W, Kamel I, Alonso R. Integrated image and speech analysis for content-based video indexing//Proceedings of the IEEE International Conference on Multimedia Systems and Computing. Hiroshima, Japan, 1996: 306-313
- [73] Han M, Hua W, Xu W, Gong Y. An integrated baseball digest system using maximum entropy method//Proceedings of the ACM Multimedia. Juan-les-Pins, France, 2002: 347-350
- [74] Nepal S, Srinivasan U, Reynolds G. Automatic detection of goal segmentations in basketball videos//Proceedings of the ACM Multimedia. Ottawa, Canada, 2001: 261-269
- [75] Assfalg J, Bertini M, Colombo C, Bimbo A. Semantic annotation of sports videos. IEEE Multimeida, 2002, 9(2): 52-60
- [76] Xie L, Kennedy L et al. Discovering meaningful multimedia patterns with audio-visual concepts and associated text. Mitsubishi Electric Research Laboratories, Technical Report 2004, 2005
- [77] Zhou W, Vellaikal A, Kuo C. Rule-based video classification system for basketball video indexing//Proceedings of the ACM Multimedia. 2000: 213-216
- [78] Vasconcelos N, Lippman A. A Bayesian framework for semantic content characterization//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Santa Barbara, USA, 1998: 566-571
- [79] Naphhade M, Huang T. Semantic video indexing using a probabilistic framework//Proceedings of the International Conference on Pattern Recognition. Barcelona, Spain, 2000: 83-88

- [80] Sun X, Jin G, Huang M, Xu G. Bayesian network based soccer video event detection and retrieval//Proceedings of the SPIE Multispectral Image Processing and Pattern Recognition. Wuhan, China, 2003: 71-76
- [81] Chang P, Han M, Gong Y. Extract highlights from baseball game video with hidden Markov models//Proceedings of the IEEE International Conference on Image Processing. New York, USA, 2002: 609-612
- [82] Wang J, Xu C, Siong C, Tian Q. Sports highlight detection from keyword sequence using HMM//Proceedings of the IEEE International Conference on Multimedia and Expo. Taiwan, 2004: 599-602
- [83] Haering N, Qian R, Sezan M. A semantic event-detection approach and its application to detection hunts in wildlife video. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(6): 857-868
- [84] Zhuang Yue-Ting, Pan Yun-He, Wu Fei. Web-Based Multimedia Information Analysis and Retrieval. Beijing: Tsinghua University Press, 2002(in Chinese)
(庄越挺, 潘云鹤, 吴飞. 网上多媒体信息分析与检索. 北京: 清华大学出版社, 2002)



TONG Xiao-Feng, born in 1976, Ph. D. . His research interests include video analysis and mining, algorithm optimization and parallelism computing.

LIU Qing-Shan, born in 1975, Ph. D. , associate profes-

sor. His current research interests include pattern recognition, machine learning and video analysis.

LU Han-Qing, born in 1961, Ph. D. , professor. His research interests include remote sensing processing, medical image system, content based image/video retrieving, objects recognition and tracking, video analysis, sport video analysis and so on.

Background

This paper gives a survey on sports video analysis, including the important ideas, algorithms, research roadmap, achievements and possible trends. This work was supported by the National Natural Science Foundation of China with title "Sports Video Analysis Based on Semantic Events" (grant No. 60475010). Sports video analysis has attracted much attention of academic and industry researchers. Its applications are somehow definite and wide-range, but involve many domains, for examples, video and audio processing, informa-

tion fusion, pattern recognition and semantics inference, communication, algorithm parallelism, etc. The National Lab of Pattern Recognition of CASIA has devoted much effort in this topic. The work involves with low-level feature extraction, middle-level key-words detection, and high-level event detection. The project team has published about 20 papers, applied for 1 patent, and 1 registration of computer software copyright.