

基于规则推理的足球视频任意球射门事件检测

华晓凤, 冯娜, 于俊清, 何云峰

引用本文

华晓凤, 冯娜, 于俊清, 何云峰. [基于规则推理的足球视频任意球射门事件检测](#) [J]. 计算机科学, 2023, 50(3): 181-190.

HUA Xiaofeng, FENG Na, YU Junqing, HE Yunfeng. [Shooting Event Detection of Free Kick in Soccer Video Based on Rule Reasoning](#) [J]. Computer Science, 2023, 50(3): 181-190.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[深度学习模型的后门攻击研究综述](#)

Backdoor Attack on Deep Learning Models:A Survey

计算机科学, 2023, 50(3): 333-350. <https://doi.org/10.11896/jsjcx.220600031>

[特征增强损失与前景注意力人群计数网络](#)

Crowd Counting Network Based on Feature Enhancement Loss and Foreground Attention

计算机科学, 2023, 50(3): 246-253. <https://doi.org/10.11896/jsjcx.220100219>

[基于深度学习的可视化仪表板生成技术研究](#)

Study on Visual Dashboard Generation Technology Based on Deep Learning

计算机科学, 2023, 50(3): 238-245. <https://doi.org/10.11896/jsjcx.230100064>

[极化自注意力约束颜色溢出的图像自动上色](#)

Polarized Self-attention Constrains Color Overflow in Automatic Coloring of Image

计算机科学, 2023, 50(3): 208-215. <https://doi.org/10.11896/jsjcx.220100149>

[基于特征融合的边缘引导乳腺超声图像分割方法](#)

Segmentation Method of Edge-guided Breast Ultrasound Images Based on Feature Fusion

计算机科学, 2023, 50(3): 199-207. <https://doi.org/10.11896/jsjcx.211200294>

基于规则推理的足球视频任意球射门事件检测

华晓凤¹ 冯 娜¹ 于俊清^{1,2} 何云峰¹

1 华中科技大学计算机科学与技术学院 武汉 430074

2 华中科技大学网络与计算中心 武汉 430074

(fengna@hust.edu.cn)

摘 要 足球视频事件检测对视频检索具有重要意义。然而,足球视频中事件较少,且主要发生在远镜头中,难以捕捉关键球员和关键动作,导致足球事件检测困难。近年来,基于深度学习的方法在足球视频事件检测上取得了一定的进展,但对事件的高层语义学习仍不够充分,检测结果有待进一步提高。如何提升足球视频事件检测的准确性是亟待解决的问题。以任意球射门事件为研究对象,提出了足球规则与深度学习相结合的事件检测模型。为了深入了解任意球射门事件的内在特性,人工总结了事件规则并在公共足球数据集上进行了验证,同时提出了规则的应用场景。针对足球视频中事件过少的问题,设计了基于规则的初始定位算法对视频进行预处理。通过多规则组合和应用,从原始视频中初步定位可能发生任意球射门事件的位置,并将其作为深度学习模型的输入进行进一步预测。在公共足球数据集上将所提模型与其他模型进行对比实验。结果表明,该模型取得了最好的效果,其精确率达到 78%,召回率达到 81.25%。相比其他模型,其精确率的提升尤为明显。可见,足球规则与深度学习相结合的任意球事件检测模型有效提升了任意球射门事件的检测性能,为足球视频中其他事件的检测提供了参考依据。

关键词: 足球视频;任意球射门;事件检测;足球规则;深度学习

中图法分类号 TP181

Shooting Event Detection of Free Kick in Soccer Video Based on Rule Reasoning

HUA Xiaofeng¹, FENG Na¹, YU Junqing^{1,2} and HE Yunfeng¹

1 School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

2 Network and Computing Center of Huazhong University of Science and Technology, Wuhan 430074, China

Abstract Soccer video event detection is of great significance to video retrieval. However, there are fewer events in soccer videos, and these events mainly occur in the far-view shot, which makes it difficult to capture key players and key actions, making soccer event detection more difficult. In recent years, methods based on deep learning have made some progress in soccer video event detection, but the learning ability of the high-level semantic of the event is still insufficient and the detection results need to be further improved. Therefore, how to improve the accuracy of soccer video event detection is an urgent problem to be solved. Taking the shooting event of free kick (free-kick shot event) as the research object, an event detection model combining soccer rules and deep learning is proposed. To have a deeper understanding of the inherent characteristics of the free-kick shot event, the event rules are manually summarized and verified on the public soccer dataset, and the corresponding application scenarios are also proposed. For the problem of too few events in soccer videos, rule-based initial localization algorithm is proposed to preprocess the videos. Through the combination and application of multiple rules, the location where the free-kick shot event may occur is initially located from the original video, which is used as the input of the deep learning model for further prediction. The proposed model is compared with other models on the public soccer dataset. Experimental results show that the proposed model achieves the best results, with the accuracy rate of 78% and the recall rate of 81.25%. Compared with other models, the improvement in accuracy is particularly prominent. It can be seen that the free-kick shot event detection model that combines soccer rules and deep learning effectively improves the performance of free-kick shot event detection and provides a basic reference for further research on the detection of other events in soccer videos.

Keywords Soccer video, Free-kick shot, Event detection, Soccer rules, Deep learning

1 引言

足球视频事件检测是体育视频分析领域的重点和难点,准确检测出足球视频中的精彩事件对于生成精彩集锦和视频

摘要、视频检索以及辅助球队分析等具有重要意义。作为“世界第一大运动”,足球运动吸引了越来越多的关注,各种足球赛事使得足球比赛视频的数量迅速增长。如何准确、快速地检测出足球视频中的精彩事件以满足检索需求是亟待解决的

问题。

足球视频事件检测面临以下挑战:1)完整的足球视频持续时间长,而精彩事件所占比例很小;2)事件发生的场景多为远镜头,球员在画面中较小且数量较多,难以获取到关键球员及其动作和球的位置;3)远镜头下不同类型事件的发生场景相似,难以判断等。基于传统机器学习的检测方法,其特征提取算法复杂,且预测事件的时序边界较为粗糙。近年来,基于深度学习的方法在足球视频事件检测上取得了一定的进展,但对事件的高层语义学习仍不够充分,检测结果有待进一步提高。

本文以直接射门的任意球事件(简称任意球射门事件)为研究对象,设计了一种足球规则与深度学习相结合的事件检测模型。利用先验知识,人工总结事件规则并将其融入事件检测模型中,以提高对事件高层语义的理解和学习。针对足球视频中事件占比较少的问题,提出了基于规则的初始定位算法。通过多规则的组合和应用,初步估计原始视频中可能发生任意球射门事件的时间点,获得候选事件片段,同时过滤大量无关片段。在此基础上,设计了基于 C3D 的任意球射门事件检测模型,对候选事件片段进行进一步处理和预测。本文提出的检测模型在 SSET 足球数据集上取得了优异的检测效果。本文的主要贡献如下:

- (1)设计了足球规则与深度学习相结合的事件检测模型,为提高对事件高层语义的学习能力,人工总结了事件规则并加以验证,将其融入事件检测模型中。
- (2)设计了基于规则的任意球射门事件初始定位模型,以排除足球视频中大量的非事件片段,获取高质量的任意球射门事件的候选事件片段集合。
- (3)设计了基于 C3D 的任意球射门事件检测模型,以进一步检测任意球射门事件及其时序边界。

2 相关工作

针对足球视频这一特定事件检测场景,研究人员深入挖掘足球视频中的高层语义信息,推动了足球视频事件检测的发展。目前,足球视频事件检测的方法可分为两类:基于传统机器学习的事件检测方法和基于深度学习的事件检测方法。

基于传统机器学习的事件检测方法结合视频、音频、文本等先验知识设计手工特征,通过多个中低层语义特征来描述不同事件的活动过程。常用的方法包括支持向量机(Support Vector Machine,SVM)^[1]、贝叶斯网络^[2-4]、隐马尔可夫模型(Hidden Markov Model,HMM)^[5-6]、基于模糊规则推理^[7]等。Wang 等结合上下文信息和足球视频的先验知识,利用动态贝叶斯网络处理时序信号,对体育视频进行建模来检测足球的精彩事件^[3]。Han 等定义了 3 类射门,采用启发式规则与无监督模糊 c 均值算法相结合的方法进行事件检测。该方法利用启发式规则过滤部分镜头,以每场比赛中的进球数作为先验知识进行模糊处理,实现进球事件检测^[7]。Xu 等提出了基于音频关键字的足球视频事件检测框架,使用 SVM 学习从低级音频特征中生成的音频关键字,应用启发式映射来检测足球视频中的语义事件^[8]。Liu 等提出了基于多特征提取与融合的事件检测方法,通过在视频流中提取视觉特征

和文本信息来检测可能包含事件的镜头,在音频流中实现了“球门”“点球”等特殊词的语音识别,并使用这些词识别事件^[9]。Hosseini 等以一组视听特征的统计信息作为输入值,通过对连续特征向量进行离散化和模糊化,形成一组元组,并构建决策树来提取元组之间的隐藏知识以及特征与事件之间的相关性;遍历决策树的所有路径,生成一组模糊规则,运用模糊推理机预测输入视频中发生的事件^[10]。Pandya 等设计了基于光流的时序边界检测的方法,以光流的变化来获取足球视频中的事件^[11]。Fakhar 等通过学习基于加权局部特征的、有监督的、有区别的、面向事件的字典,将足球视频事件检测任务定义为稀疏表示问题^[12]。Morra 等提出了检测原子和复杂事件的两层系统。原子事件是基于被检测对象的时间和逻辑组合、它们的相对距离,以及速度和加速度等时空特征来检测的。复杂事件被定义为原子事件和复杂事件的时序和逻辑组合,并用间隔时序逻辑(Interval Timing Logic,ITL)表示,通过 ITL 执行推理任务^[13]。

基于深度学习的方法通过对大规模样本数据的训练学习来检测事件,提高了模型的泛化性^[14-16]。Jiang 等将卷积神经网络(Convolutional Neural Network,CNN)和循环神经网络(Recurrent Neural Network,RNN)相结合,构建了一种用于足球视频事件检测的深度神经网络。处理过程分为 3 个阶段:第一阶段采用传统方法确定足球视频事件边界,将视频划分为若干个 play-break 片段;第二阶段利用预先训练好的 CNN 从 play-break 片段中提取关键帧的语义特征;第三阶段利用 RNN 将从 play-break 片段中提取的语义特征映射到足球事件类型^[17]。Song 等提取镜头中的视频帧序列和光流序列,利用 I3D 网络对视频单元进行分类,输出每个视频单元的预测概率值;在分类的基础上,利用基于概率值的分组方法和基于类的分组方法将相邻的片段组合在一起以定位事件的边界,最终使得事件检测的 mAP 提高到 62%^[18]。Feng 等对比了 VGGNet,GoogleNet,ResNet 和 C3D 网络在足球视频事件检测上的效果。其中,前 3 种网络模型提取单个视频帧的空间特征,添加 LSTM 层将单帧特征融合来描述视频片段;第四种模型可同时提取视频片段的时空特征。实验中使用滑窗法生成候选片段,利用上述网络模型预测不同类型的事件。实验结果表明 C3D 取得了最好的检测效果^[19]。

现有的足球视频事件检测方法不能满足实际的检测需求。基于传统机器学习的检测方法对足球事件的检测较为准确,但检测的事件类型有限且预测的事件的时序边界粗糙。基于深度学习的检测方法对事件的高层语义信息学习不充分,且对小样本事件的检测性能较差。为了进一步提高足球视频事件的检测效果,本文以任意球射门事件为研究对象,提出足球规则与深度学习相结合的事件检测模型,以探究事件规则对事件检测的作用效果。

3 任意球射门事件规则

足球运动的竞赛规则和足球视频的剪辑方式对于事件的呈现具有一定的规律,可作为先验知识应用于事件检测中。本节对任意球射门事件进行深入分析,人工总结事件规则并进行验证,同时分析了规则的应用场景。

3.1 规则描述

在足球比赛中,对任意球射门事件的判罚有明确的规则,导播也会采取一定的拍摄手法向观众展示完整的事件过程,这些规则和拍摄手法呈现一定的规律,表达了任意球射门事件的特性。通过对足球竞赛规则的学习和对足球视频中任意球射门事件样本的观察,从视觉、音频和文本等方面归纳了任意球射门事件的规则。

规则 1 任意球射门事件发生在中远镜头。任意球射门事件只可能发生在中镜头和远镜头中,有且只有两种镜头结构:单一远镜头结构或由中镜头切换至远镜头(以“中-远镜头”表示),如图 1 所示。在中-远镜头中,导播通常先切换至中近镜头来呈现主罚球员的发球状态,球发出后,再迅速切换到远镜头。

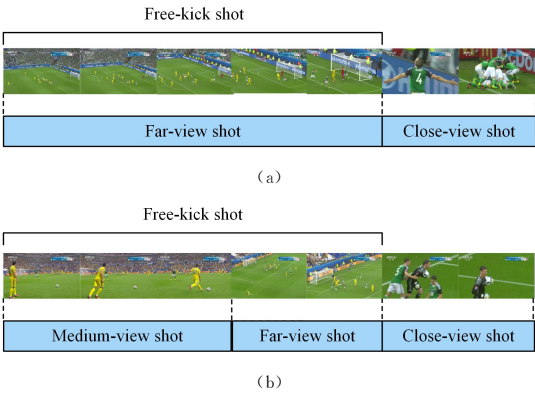


图 1 任意球射门事件的两种镜头结构

Fig.1 Two shot structures of free-kick shot event

规则 2 任意球射门事件主体发生在远镜头的初始阶段。由规则 1 可知,任意球射门事件的主要部分发生在远镜头中,且为远镜头的第一个事件。然而,初始阶段的持续时间并不统一,一些事件在切换至远镜头后立刻发生,一些事件则会用较长时间来展现全场的准备状态。

规则 3 任意球射门事件发生在球场两端。任意球射门的发球位置是犯规行为发生的位置或者发生犯规时足球的位置,主罚球员的站位可能在罚球区前方、罚球区两侧或者球场边线附近等,如图 2 所示。

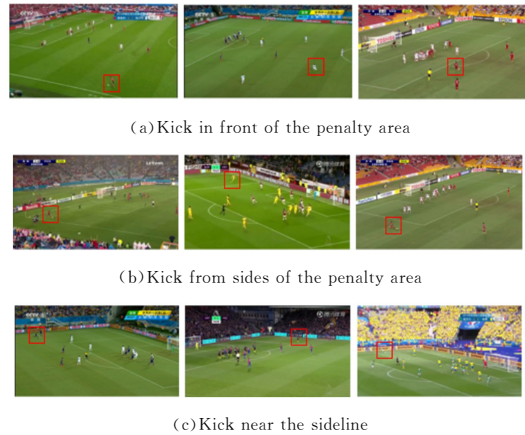


图 2 任意球射门事件的不同发球位置

Fig.2 Different kick positions of free-kick shot event

规则 4 任意球射门事件发生时,比赛处于中断状态。任意球射门是足球比赛中发生犯规后重新开始比赛的一种方法。犯规事件发生时,裁判吹哨示意比赛暂停,进而判罚任意球。

规则 5 任意球射门事件具有特定的运动特征:发球前,球员聚集在球门附近等待发球;发球后,球员向球门方向移动。

规则 6 部分任意球射门事件在其所在的远镜头中有较长时间的全场准备状态,如裁判划定罚球点、球员形成防守站位等。

规则 7 部分任意球射门事件发生前有防守球员形成的严密人墙。人墙是任意球射门事件的典型特征,如图 3 所示,球员在球门前站成一排,以阻挡进球。然而多数情况下,球员聚集在罚球区附近,站位较为随意。



图 3 任意球射门事件中防守球员站位

Fig.3 Position of defenders in free-kick shot event

规则 8 部分任意球射门事件发生前有犯规回放。

规则 9 部分精彩的任意球射门事件发生后有相应的事件回放。

规则 10 部分任意球射门事件发球时有哨声示意。

规则 11 部分任意球射门事件发生时“任意球”关键字解说。

规则 12 直播文本中有对任意球事件的说明。

在以上规则中,规则 1—9 从足球视频的视觉内容、运动状态、镜头结构、视频剪辑方式等方面进行总结,规则 10—12 从音频信号、音频关键字和文本关键字等方面进行总结。其中,规则 1—5 是所有的任意球射门事件的共同特性,而规则 6—9 说明了部分任意球射门事件的特点。

3.2 规则验证

任意球射门事件规则是人工观察样本总结得到的,无法说明具体的分布情况,可能存在一定偏差。为了验证事件规则的有效性,我们在小样本数据集上对规则进行数据统计和分析,保留有效的规则。从 SSET 足球视频数据集^[19]中随机选取不同赛事、不同天气、不同光照条件、不同地面条件的 20 个足球视频构建验证集,共包含 48 个任意球射门事件,统计结果如下。

(1)任意球射门事件不同镜头结构统计结果如表 1 所列,87.5%的任意球射门事件在一个远镜头中完成发球、球在空中运动、球落地或被其他球员接住的完整过程,只有 12.5%的任意球射门事件由中远镜头共同展现其完整过程。这表明任意球射门事件的主体在远镜头中呈现,由此验证了规则 1 和规则 2 的有效性。

表 1 镜头结构分布情况

Table 1 Distribution of shot structures

Shot structure	Number of events	Proportion/%
Far shot	42	87.5
Medium-far shot	6	12.5

(2)任意球射门事件的罚球位置统计结果如表 2 所列,70.84%的任意球射门事件在罚球区前方发球,其他两类占比均为 14.58%。其中,罚球点在罚球区两侧的任意球射门事件存在一类特殊的情况,即罚球位置非常接近角球区,由于拍摄角度和运动方式与角球事件极为相似,因此易与角球事件混淆。

表 2 不同踢球位置分布情况

Table 2 Distribution of different kick positions

Kick position	Number of events	Proportion/%
In front of the penalty area	34	70.84
Sides of the penalty area	7	14.58
Near the sideline	7	14.58

(3)比赛发生中断的重要特征是镜头切换。足球视频剪辑时,任意球射门事件发生前会多次切换镜头来呈现准备状态。描述准备状态的镜头类型包括远、中、近 3 类,镜头数量通常在 2 个以上,镜头的持续时间存在较大差异。重点分析远镜头,如表 3 所列,约有 90%的任意球射门事件在准备状态中包含远镜头,但远镜头的数量较为随机,无明显规律。描述准备状态的远镜头共有 63 个,其持续时间分布如表 4 所列。这些远镜头的持续时间大多集中在 50~300 帧的范围内,最短只有 31 帧,最长为 578 帧。仅通过持续时间难以区分准备状态下的远镜头与发生任意球射门事件的远镜头。此外,部分切换的远镜头并不预示着比赛的中断,例如在球员持球时,会从远镜头切换到中镜头,对持球球员进行特写,再切回远镜头。这些镜头构成远-中-远的镜头结构,且第二个远镜头中不可能发生任意球事件。因此,进一步考虑通过排除具有远-中-远镜头结构的第二个远镜头来判断比赛是否中断。经统计,有 17 个任意球射门事件所在的远镜头符合远-中-远的镜头结构,占比为 35.42%。由此可见,运用该规则进行排除时,部分任意球射门事件也会被过滤掉。目前尚未找到有效方法能准确检测比赛中断。

表 3 准备状态下远镜头的分布情况

Table 3 Distribution of far-view shot in ready state

Number of far-view shots in ready state	Number of events	Proportion/%
0	5	10.42
1	28	58.33
2	10	20.83
大于等于 3	5	10.42

表 4 准备状态下远镜头持续时间分布情况

Table 4 Distribution of duration of far-view shot in ready state

Duration/frame	Number of events	Proportion/%
小于等于 50	3	4.76
50~100	16	25.40
100~200	27	42.86
200~300	13	20.63
大于等于 300	4	6.34

(4)对规则 6 进行验证。经过统计,有 12 个任意球射门事件,其所在的远镜头中全场准备状态的持续时间超过 100 帧,所占比例为 25%,其中最长持续时间达到了 370 帧。

(5)在任意球射门事件中,形成人墙的事件所占比例为

22.91%。但严密人墙检测难度大,且只能覆盖小部分的任意球射门事件。

(6)犯规回放和任意球射门回放的统计结果如表 5 所列,其中犯规回放共有 29 个,任意球射门回放共有 19 个。在任意球射门事件中,事件前后均无回放和事件前后均有回放所占的比例皆约为 1/3。

表 5 回放分布情况

Table 5 Distribution of replays

Situations	Number of events	Proportion/%
No foul replay & no free-kick shot replay	16	33.33
Foul replay only	13	27.08
Free-kick shot replay only	3	6.25
Foul replay & free-kick shot replay	16	33.33

3.3 规则应用

3.2 节统计了与规则有关的数据,并在数据集上进行了验证。本节的重点是考虑如何将事件规则融入事件检测模型当中。

现有的事件检测方法对足球视频的高层语义信息的学习能力有限,检测效果并不理想。而人工总结的规则包含丰富的事件语义信息,可以应用于任意球射门事件的初始定位和事件校验工作中。在事件的初始定位中,应用多条规则对一个完整的足球视频进行筛选,过滤无关片段,能够减少候选片段的数量,提高候选片段的质量,同时降低后续的计算成本。在事件校验中,应用简单的规则去除部分预测错误的视频片段或增加部分符合任意球射门事件规则的视频片段,能够提升事件检测的精确率和召回率。由此可见,人工总结的规则对任意球射门事件的检测具有重要意义和价值。

将事件规则加入事件检测模型中时,需要综合考虑特征的检测成本、对非同类事件的区分性以及对同类事件的覆盖比例。经分析,人工总结的事件规则可应用于多种场景。规则 1—4 是任意球射门事件的共同特性,可应用于事件的初始定位,得到可能发生任意球射门事件的视频片段。规则 5 描述的是事件的运动特征,可利用深度学习提取视频片段的时空特征以区分不同的事件类型。规则 6—9 可用于事件校验,即在预测事件时,增大符合该规则的预测事件的置信度。

4 任意球射门事件检测

4.1 模型设计

本文提出了规则推理与深度学习相结合的任意球射门事件检测模型,分为两个模块:基于规则的任意球射门事件初始定位和基于 C3D 的任意球射门事件预测。前者输出可能包含任意球射门事件的候选视频片段,后者在此基础上进一步预测任意球射门事件及其时序边界,整个流程如图 4 所示。在初始定位模块中,主要应用了两条规则:任意球射门事件的主体发生在远镜头的初始阶段(图中记为①)和任意球射门事件发生在球场两端(图中记为②)。包含以下任务:镜头边界检测和镜头分类、球场两端检测和生成候选事件片段。在预测模块中,利用 C3D 模型提取视频片段特征并预测事件。

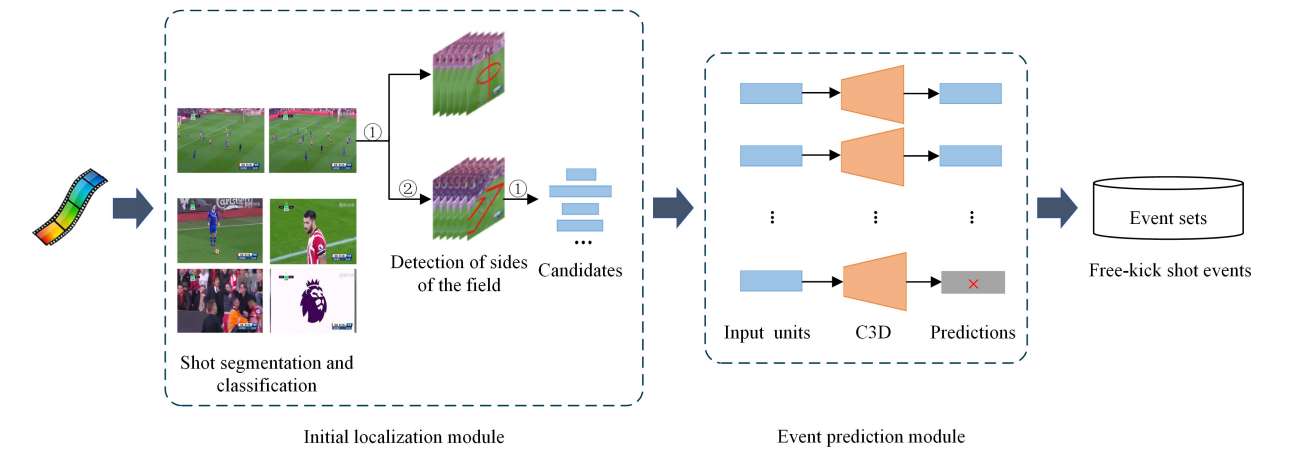


图 4 任意球射门事件检测模型

Fig. 4 Free-kick shot event detection model

4.2 基于规则的任意球射门事件初始定位

4.2.1 镜头边界检测

目前已经有了大量的镜头边界检测算法并取得了不错的检测效果。本文采用 Wang 等提出的镜头边界检测算法,该方法分为 3 个阶段:第一阶段生成候选边界帧,利用 CNN 提取视频帧特征并计算相邻帧的差异度,得到候选边界帧;第二阶段检测切变,基于 3D CNN 提取候选边界帧附近片段的时空特征,识别出切变帧、渐变候选帧与非边界帧;第三阶段检测渐变,使用 3D CNN 分别预测候选帧为渐变开始、中间和结束的概率,基于这 3 种概率定位镜头渐变的起止时间,得到渐变序列^[20]。在足球视频中,渐变主要发生在 logo 序列,因此在第二和第三阶段,利用 SSET 足球数据集的 logo 序列样本对网络进行微调,以提高对 logo 序列的检测效果。

4.2.2 镜头分类

按不同的语义可将足球视频的镜头划分为 5 类:远镜头、中镜头、近镜头、场外镜头以及回放镜头。其中,回放镜头可以通过 logo 序列匹配得到。

目前,镜头分类算法主要基于主色比率、球员的大小和数量等传统特征实现,计算复杂,中、远镜头容易混淆,分类效果有待进一步提升。近年来,深度学习在图像分类上取得了优异的表现,本文基于 VGG^[21] 网络构建了镜头分类模型,如图 5 所示。

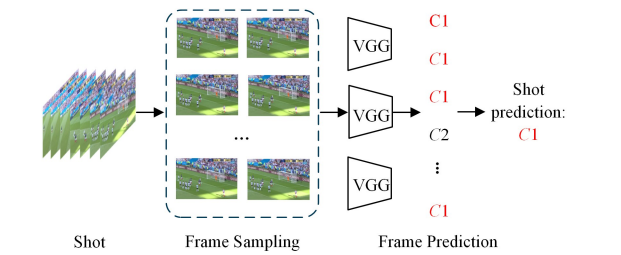


图 5 镜头分类模型

Fig. 5 Shot classification model

在一个镜头中随机采样固定数量的视频帧,分别输入 VGG 网络中提取帧特征并预测该帧的镜头类型。统计一个镜头中被预测为不同镜头类型的帧数,选取帧数最多的镜头

类型作为该镜头的预测类型。根据镜头分类的结果,保留所有的远镜头,进行下一步处理。

4.2.3 球场两端检测

根据规则“任意球射门事件发生在球场两端”,需要进一步检测远镜头中的球场两端片段。按场景区域划分,远镜头可以分为球场两端区域和中场区域。球场两端区域有球门、罚球区、角球区等典型特征,但由于光照阴影、球员遮挡、拍摄的角度不同等因素的干扰,因此检测难度有所增加。而在中场区域,有中线、中圈、长边线等典型特征,这些特征区分性强,不易受噪声影响,因此可将球场两端区域的检测转换成中场区域的检测。常用的中场检测方法是利用中圈检测实现的,但中圈检测算法时间复杂度较高、鲁棒性较差,而中线易检测、噪声小,因此选择中线检测来实现中场区域的识别,检测流程如图 6 所示。

(1) 球场区域分割。将视频帧转换到 HSV 颜色空间,运用黄色和绿色的 HSV 范围检查每个像素点,得到二值化图像。利用 openCV 中的腐蚀和膨胀函数,去除球场球员和标志线;利用 openCV 的孔洞填充算法,填补二值图中的细小孔洞,并检测图中轮廓;将面积较小的连通区块用白色填充,得到较为完整的球场区域和场外区域,将得到的图像记为 *img_color*。

(2) Tophat 处理。将视频帧转换为灰度图,用白色像素替代小于阈值的像素点,将得到的图像记为 *img_tophat*。

(3) 获取球场区域的标志线。将 *img_color* 和 *img_tophat* 做“与”操作,即可获取球场区域的标志线,记为 *img_merge*。

(4) 直线检测。运用霍夫变换,对 *img_merge* 进行线段检测和拟合,得到候选线段集合 $L: \{L_1, \dots, L_i, \dots, L_n\}$ 。实验中,将倾角在 $90^\circ \pm 15^\circ$ 范围内的线段加入竖直线段集合,并计算线段的相关参数。

(5) 中场判定。在足球视频中,中线通常是一条垂直于球场长边线、竖直或近似竖直的白色线段。本文限制中线的位置在 $5 \sim 630\text{px}$ 范围内。在竖直线段集合中,若存在一条满足位置要求的线段,则判定该线段为中线,对应的视频帧即为中场区域。

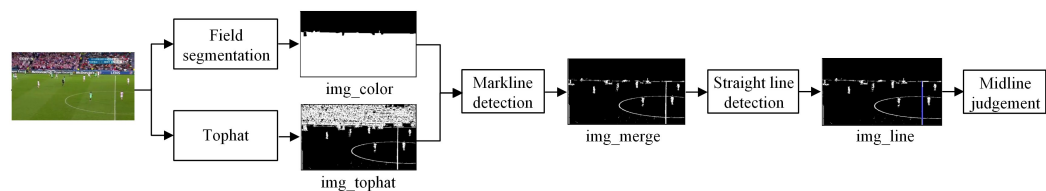


图6 中场检测流程

Fig. 6 Process of midfield detection

基于中场区域识别的结果,进一步确定球场两端片段的时序边界。从远镜头的第一帧开始,判定当前帧是否为中场区域。若不是中场区域,则反向推断为球场两端区域,按一定的间隔继续向后采样;若是中场区域,则停止对该远镜头的采样,当前帧即为球场两端片段的结束帧。该远镜头的起始帧与结束帧及其之间的视频帧构成完整的球场两端片段。

考虑到中场区域识别错误可能会导致球场两端片段的漏检或长度过短,因此增加了容错性处理,具体如下:若当前帧被识别为中场区域,分别隔10帧、20帧各采样一帧图像进行识别;若这两帧均被识别为中场区域,则认定当前帧是球场两端片段的结束帧;否则,继续向后采样。

4.2.4 初始定位

一个远镜头可能存在多个球场两端片段,且持续时长不一。因此,结合规则“任意球射门事件主体发生在远镜头的初始阶段”,以远镜头中第一个持续时间大于阈值的球场两端片段作为候选事件片段,可直接过滤该远镜头的其余片段,减少计算量。通过以上操作获取可能发生任意球射门事件的视频片段,但得到的候选事件片段的位置和时序边界较为粗糙,仍需对任意球射门事件进行更准确的定位。

4.3 基于C3D的任意球射门事件检测

通过初始定位得到的任意球射门事件的候选片段持续时间较长,且仍存在很多非任意球事件片段,如角球、界外球等,需要进一步处理,同时细化事件发生的时序边界。本文基于C3D^[22]构建了任意球射门事件预测模块,如图7所示。

利用滑动窗口对每个候选事件片段生成预测单元,作为C3D模型的输入来提取时空特征并预测分类。被预测为任意球射门事件的预测单元经过进一步筛选、合并,最终输出完整的任意球射门事件及其时序边界。

4.3.1 预测单元生成

由初始定位模型生成的候选事件片段的长度较长,且差异较大。此外,任意球射门事件的平均持续时长为50帧,明显小于候选事件片段的长度。因此,本文采用滑窗法对单个候选片段生成若干个预测单元。设置固定大小的滑动窗口和滑动步长,从每个候选事件片段的起始帧开始向后滑动,滑动窗口内的视频帧即为一个预测单元,按照滑动步长继续滑行直至候选事件片段结束。当滑动窗口移动到候选事件片段的末尾,且滑动窗口内的帧数小于滑窗的长度时,向反方向移动滑动窗口,直至达到滑动窗口的长度,生成该候选事件片段的最后一个预测单元。

4.3.2 时序边界检测

利用C3D模型学习任意球射门事件的表现特征和运动模式,实现任意球射门事件的准确检测和定位。将预测单元分别输入C3D网络中提取时空特征,得到每个预测单元的概率值。考虑到视频相邻帧之间存在较高的相似性,无须对预测单元中的每一帧进行计算。因此,本文对每个预测单元进行隔帧采样,将实际计算的预测单元的长度缩短至预测单元长度的一半,有效减少了每个预测单元的计算量。

4.3.3 后处理

对预测结果进行后处理,确定任意球射门事件及其时序边界,具体如下:过滤预测为非事件的预测单元;筛选任意球射门事件的概率值大于阈值的预测单元;将相邻的预测单元合并,作为一个任意球射门事件。最终输出足球视频中所有的任意球射门事件。与候选事件片段相比,该模型输出的任意球射门事件的时序边界更精准。

5 实验

5.1 数据集

采用SSET足球视频数据集进行实验,随机抽取201个视频作为训练集,20个视频作为测试集。测试时,对视频执行完整的任意球射门事件检测算法。

在SSET数据集中,对任意球射门事件的时序边界标注过于严格,缺失了与事件有关的上下文信息,如任意球射门前球员的准备状态。因此,本文对训练集中的任意球射门事件进行扩充,增加上下文信息。扩充方法如算法1所示。

算法1 事件片段扩充算法

输入:任意球射门事件的原始时序边界(start,end),滑动窗口的长度L
输出:任意球射门事件扩充后的时序边界(new_start,new_end)
1. 计算任意球射门事件的持续时间 duration。

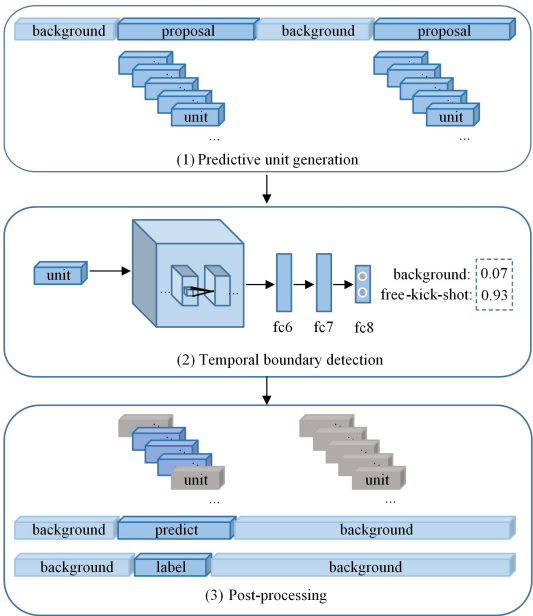


图7 任意球射门事件预测模块

Fig. 7 Free-kick shot event prediction module

2. 若 $\text{duration} < L * 3/4$, 则分别向左、向右扩展 deta 帧图像, 保证原始事件与扩展后的事件的 $\text{IoU} > 0.7$, 即 $\frac{\text{duration}}{(\text{duration} + 2 * \text{deta})} > 0.7$ 。扩展后的事件的时序边界为 $(\text{start}', \text{end}')$ 。

3. 向左扩展 $L/4$ 帧图像, 得到事件的时序边界为 $(\text{start}' - L/4, \text{end}')$ 。

配置正负样本训练集时, 所有任意球射门事件为正样本集, 候选事件片段中不包含任意球射门事件的片段为负样本集。考虑到任意球射门事件的数量较少, 且与角球难以区分, 因此, 训练时正负样本比例设为 1:2。其中, 负样本中角球与其他负样本的比例为 1:1。网络模型训练时, 对正负样本随机采样一个滑窗大小的视频帧序列。

5.2 网络模型设置

在镜头分类模型中, 采用 VGG-16。训练时, 加载预训练的模型进行微调。损失函数由分类损失和 L2 正则化组成, L2 正则化衰减系数为 5×10^{-4} 。初始学习率为 0.001, 当损失不再下降时, 将学习率下调至 0.0001 继续训练至损失不再下降。采用随机梯度下降法更新网络参数。

在事件预测模型中, 滑动窗口大小为 64 帧, 滑动步长为 16 帧, C3D 模型输入的视频序列长度为 32 帧。训练时, 加载预训练的模型进行微调。损失函数由分类损失和 L2 正则化组成, L2 正则化衰减系数为 5×10^{-4} 。初始学习率为 0.001, 当损失不再下降时, 将学习率下调至 0.0001 继续训练至损失不再下降。 dropout 为 0.5。采用随机梯度下降法更新网络参数。

5.3 评价指标

采用精确率 (precision)、召回率 (recall) 和 F1-measure 作为评价指标来评估任意球射门事件检测模型的检测效果, 计算式如式 (1) — 式 (3) 所示:

$$\text{precision} = \frac{\text{correct}}{\text{correct} + \text{false}}$$

(1)

$$\text{recall} = \frac{\text{correct}}{\text{correct} + \text{miss}}$$

(2)

$$\text{F1} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

(3)

5.4 初始定位实验结果与分析

本节在 221 个完整的足球视频上进行任意球射门事件的初始定位, 为后续的任意球射门事件的训练和测试提供候选事件片段。

5.4.1 镜头边界检测

镜头边界检测结果如表 6 所列。

表 6 镜头边界检测结果
Table 6 Results of shot boundary detection
(单位: %)

Shot transition type	Precision	Recall	F1-measure
Abrupt	98.3	99.6	98.9
Logo sequence	99.6	99.9	99.7

其中, 镜头切变和 logo 序列的召回率均在 99.6% 以上。镜头切变的误差主要发生在中、近镜头, 不影响远镜头的分割。

5.4.2 镜头分类

镜头分类的实验结果如表 7 所列。该模型对镜头的识别效果较好, 其中远镜头的准确率最高, 达到了 99.7%, 召回率为 98%。

表 7 镜头分类结果
Table 7 Results of shot classification
(单位: %)

Shot type	Precision	Recall	Total precision
Far-view	99.77	98.04	96.60
Medium-view	91.58	98.58	
Close-view	97.95	96.40	
Out-of-field-view	96.07	89.80	

镜头分类的误差分为 4 类: 远镜头被误判为中镜头; 中镜头被误判为近镜头和场外镜头; 近镜头被误判为中镜头和场外镜头; 场外镜头被误判为中镜头和近镜头, 极少数被误判为远镜头。造成中远镜头误判的主要原因是一些镜头由远镜头拉到中镜头或者由中镜头推到远镜头, 但这类镜头通常不会发生任意球射门事件。而其他镜头的分类错误也几乎不会影响任意球射门事件的初始定位和检测。

不同方法在足球视频数据集上的镜头分类效果如表 8 所列。Wu 等提出了基于波动检测的镜头分类方法, 球场像素比例在阈值附近发生变化则为一次波动, 该方法按波动次数区分中镜头和远镜头^[23]。Cheng 等根据预定义的类别对各个镜头进行标记, 再利用 3D CNN 学习时空特征, 对采样的视频片段进行分类^[24]。Sarkar 等提出双流深度神经网络进行镜头分类, 将一条分支输入 RGB 图像, 另一条分支输入视频帧中的球员属性, 两个特征经过融合后输入模型中进行镜头分类^[25]。

由表 8 可知, 不同方法对远、中、近镜头的识别能力存在较大差异。从整体来看, 本文设计的分类模型表现优于其他方法, 对 3 类镜头的召回率均在 96% 以上。Wu 等利用传统方法进行镜头分类, 对近镜头的分类效果较好, 但对中、远镜头的识别效果并不理想。Cheng 等利用 3D CNN 捕获镜头的时空特征, 提升了镜头分类的性能, 但对远镜头的识别能力较差。Sarkar 等提出的方法对中镜头和近镜头的识别效果并不理想, 但在远镜头的识别上达到了最好的效果, 表明球员信息的融入对远镜头的识别性能有明显的提升。但该方法需要利用 faster R-CNN 检测球员, 并计算球员的相关属性, 复杂度较高, 并且对中镜头和近镜头的分类效果不如其他方法。

表 8 不同方法的镜头分类结果
Table 8 Shot classification results of different methods
(单位: %)

Shot type	Wu 等 ^[23]		Cheng 等 ^[24]		Sarkar 等 ^[25]		Our method	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Far-view	89.0	93.8	95.4	88.2	98.0	100.0	99.8	98.0
Medium-view	92.6	88.0	93.6	90.1	93.0	80.0	91.6	98.6
Close-view	99.7	98.9	96.8	92.8	82.9	92.0	98.0	96.4

5.4.3 初始定位

为了更好地评估事件规则在初始定位上的应用效果,本文采用以下评价指标:候选事件片段数量占镜头总数的比例和候选事件片段的总时长占视频总时长的比例,统计结果如表 9 所列。可以看出,通过镜头分割和识别得到的远镜头的总持续时间只占完整视频的 61.87%,远镜头数量只占镜头总数的 32.91%。仅用规则 1,就能去除大量的中、近、场外和回放镜头。在远镜头的基础上进一步筛选,得到候选事件片段集合,其总时长只占视频总时长的 15.35%,数量只占镜头总数的 16.64%。这表明利用规则 2 和规则 3 能够在远镜头上初步定位出任意球射门事件的范围,进一步减少候选事件片段的数量和持续时间。基于规则的初始定位,在减少了后续计算量的同时提高了事件检测的精确率。在候选事件片段中包含了所有的任意球射门事件,这说明本文提出的多规则应用的方案是有效的。此外,候选事件片段中也包含了角球、点球、界外球、球门球以及发生在罚球区附近的其他视频片段。其中,角球的数量较多,容易与任意球射门发生误判,从而影响任意球射门事件的检测效果。因此,在后续的工作中,需要对这两类事件设计有效区分的算法。

表 9 候选事件片段数量和时长统计
Table 9 Statistics of number and duration of proposal event segments

Type	Number of segments	Proportion of segments/%	Total duration	Ratio of total duration to video duration / %
Candidate	11 898	16.64	2 434 107	15.35
Far-view shot	23 528	32.91	9 813 317	61.87
Video	71 498	100.00	15 861 717	100.00

5.5 事件检测实验结果与分析

针对测试集中的每个视频,分别执行完整的任意球射门事件检测流程,实验结果如表 10 所列。任意球射门事件的精确率和召回率分别为 78%和 81.25%,且两者的差距不大。这表明本文方法对任意球射门事件检测是有效的。

表 10 任意球射门事件检测结果
Table 10 Detection results of free-kick shot event
(单位:%)

Precision	Recall	F1-measure
78.00	81.25	79.59

预测的事件片段包含对应的完整真实片段,但长于真实片段,涵盖了任意球射门事件部分上下文信息。对误差进行进一步分析,误差包括错检和漏检,其中错检的样本为 11 个,漏检样本为 9 个。错检主要是将角球和间接任意球误检为任意球射门。出现错检的原因是这两类事件与任意球射门事件在外观和运动特征方面比较相似,容易被错检为任意球射门事件。另一类错检为同一个候选事件片段中检测出两个任意球射门事件,可以利用规则进行校验,即一个远镜头只可能发生一个任意球射门事件且发生在远镜头的初始阶段,由此对第三类错检进行纠正,进一步提高精确率。漏检样本中有 6 个任意球射门事件发生在罚球区两侧区域,其中 3 个靠近角球区。由于这类样本数量少,C3D 模型不能充分学习该类事件的特征。以上分析表明,本文构建的检测模型需要进一步

提升对任意球射门事件的时空特征表达能力,以区分相似事件。

不同方法对任意球射门事件的检测结果如表 11 所列。Yu 等利用 VGG 提取视频帧特征,对当前帧及其之前 L 个视频帧的特征进行重组,即每一个时间节点的输入为多帧特征的加权和,应用 LSTM 传递视频帧之间的依赖关系。此外,还利用回放对事件预测进行了增强^[26]。Vanderplaetse 等使用滑窗法生成候选片段,提取每个候选片段的视频流和音频流特征,将其维度降至 512 维;之后采用浅层池化神经网络,在 dropout 层将两类特征融合,输出候选事件片段的预测分类;经过后处理,得到各类事件的预测片段^[27]。本文还针对规则进行了消融实验,其中,C3D 的训练和测试都基于完整的足球视频,即将滑动窗口直接作用在完整的足球视频中,生成待检测的事件片段进行预测。C3D 结合规则 2(任意球射门事件发生在球场两端)首先将视频进行镜头分割和识别,保留远镜头进行球场两端片段的检测,将输出的球场两端片段作为候选事件片段进行预测。

表 11 不同方法的事件检测结果
Table 11 Event detection results of different methods
(单位:%)

Methods	Precision	Recall	F1-measure
Yu 等 ^[26]	5.40	72.90	10.10
Vanderplaetse 等 ^[27]	4.05	64.89	7.62
C3D ^[22]	21.10	83.33	33.80
C3D & rule 2	29.10	81.25	42.80
C3D & rule 1 & rule 2	78.00	81.25	79.59

从表 11 可知,本文方法对任意球射门事件的检测性能明显优于其他方法,其精确率和召回率都取得了不错的效果,尤其是精确率有明显的提升。这说明本文提出的基于规则的初始定位在提升事件检测的精确率方面发挥着重要作用。Yu 等提出的模型对任意球射门事件检测的召回率较高,但精确率只有 5.4%,远低于后 3 种方法。该方法选用 VGG 提取视频帧特征,进而用 LSTM 对帧特征进行融合,而后 3 种方法采用 C3D 直接提取视频片段的时空特征,这表明 C3D 对事件特定的运动模式和内在规律的学习能力更强,在事件检测任务上表现更好。精确率产生巨大差距的主要原因是 Yu 等是在完整的足球视频上检测事件,将大量的无关片段预测为任意球射门事件,因此造成错检。Vanderplaetse 等提出的模型的检测效果最差,F1-measure 只有 7.62%,其中召回率为 4.25%,可能的原因是任意球射门事件的音频流特征并不明显,对任意球射门事件的检测性能提升没有帮助。从后 3 行结果可得规则的应用对事件的检测有明显提升。在 C3D 方法中,由于训练和测试都基于完整的足球视频,与其他诸如射门、回放中的任意球射门的情况难以区分。增加规则 2 后,首先排除了回放镜头,其次,训练样本只需要从球场两端的视频片段中生成,训练更具有针对性,进一步提高了模型的识别能力。增加规则 1(任意球射门事件发生在远镜头的初始阶段)后,对球场两端片段进行筛选,排除了大部分的射门、传球等容易与任意球射门发生误判的视频片段,使得模型对任意球射门事件的识别更加准确。

表 12 列出了不同方法处理单个视频的平均时长。可以看出,Vanderplaetse 等提出的方法速度最快,而 C3D 模型处理的速度最慢。增加规则后,尽管增加了镜头分割与镜头分类和球场两端检测算法,但同时也减少了 C3D 模型的输入。从最终结果来看,增加规则后的处理速度比 C3D 更快,这表明应用规则在提升效果的同时也提高了计算效率。规则 1、规则 2 与 C3D 模型的结合能更好地平衡事件检测结果和所需时长的关系,达到最优效果。

表 12 不同方法处理单个视频的平均时长
Table 12 Average duration to process a single video by different methods

Methods	Average duration /min
Yu 等 ^[26]	50
Vanderplaetse 等 ^[27]	20
C3D ^[22]	105
C3D & rule 2	75
C3D & rule 1 & rule 2	45

结束语 本文针对足球比赛视频,以任意球射门事件为研究对象,提出将人工总结的事件规则与深度学习相结合来构建任意球射门事件检测模型;通过对任意球射门事件的深入分析,总结了 12 条事件规则,并在 SSET 足球数据集上进行验证,同时分析了规则的应用场景;提出了基于规则的初始定位算法,对原始视频执行镜头边界检测、镜头分类以及初始定位等预处理操作,以生成高质量的任意球射门事件的候选片段集合;采用 C3D 模型进一步对候选事件片段进行处理和预测,以获得更加精准的任意球射门事件及其时序边界。实验结果表明,所提模型优于其他算法,提升了任意球射门事件的检测性能。然而本文方法中采用的球场两端片段检测算法是基于线检测实现的,容易受到球场环境,如光照、雾气等因素的干扰,存在误差。此外,事件检测算法中任意球射门事件与部分角球事件存在相似性,容易发生误判。如何有效判别任意球射门事件和角球事件,还有待进一步研究。

在未来的工作中将继续优化本文算法,提高算法的鲁棒性,并将其进一步推广到其他足球事件检测中。

参 考 文 献

[1] YE Q,HUANG Q,GAO W,et al. Exciting event detection in broadcast soccer video with mid-level description and incremental learning[C]//Proceedings of the 13th Annual ACM International Conference on Multimedia. New York: ACM, 2005: 455-458.

[2] TONG X,LU H,LIU Q. A three-layer event detection framework and its application in soccer video[C]//Proceedings of the International Conference on Multimedia & Expo. Piscataway, NJ:IEEE,2004:1551-1554.

[3] WANG F,MA Y F,ZHANG H,et al. Dynamic bayesian network based event detection for soccer highlight extraction[C]//Proceedings of the 2004 International Conference on Image Processing. Singapore:IEEE,2004:633-636.

[4] XU H,FONG T H,CHUA T S. Fusion of Multiple Asynchronous Information Sources for Event Detection in Soccer

Video[C]//Proceedings of the 2005 IEEE International Conference on Multimedia and Expo. Piscataway, NJ: IEEE, 2005: 1242-1245.

[5] XIONG Z,RADHAKRISHNAN R,DIVAKARAN A,et al. Audio events detection based highlights extraction from baseball, golf and soccer games in a unified framework[C]//Proceedings of the 2003 International Conference on Multimedia and Expo. Piscataway,NJ:IEEE,2003:401-404.

[6] JIN G,TAO L,XU G. Hidden Markov Model Based Events Detection in Soccer Video[C]// Proceedings of the International Conference Image Analysis and Recognition. Berlin, German: Springer. 2004:605-612.

[7] HAN Y,LIU G,CHOLLET G. Goal event detection in broadcast soccer videos by combining heuristic rules with unsupervised fuzzy c-means algorithm[C]//Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision. Hanoi, Vietnam:IEEE,2008:888-891.

[8] XU M,MADDAGE N C,XU C,et al. Creating audio keywords for event detection in soccer video[C]//Proceedings of the 2003 International Conference on Multimedia and Expo. Piscataway, NJ:IEEE,2003:281-284.

[9] LIU H,HE T. Integrating Multiple Feature Fusion for Semantic Event Detection in Soccer Video[C]//Proceedings of the 2009 International Joint Conference on Artificial Intelligence. Hainan,China:IEEE,2009:128-131.

[10] HOSSEINI M S,EFTEKHARI-MOGHADAM A M. Fuzzy rule-based reasoning approach for event detection and annotation of broadcast soccer video [J]. Applied Soft Computing, 2013,13(2):846-866.

[11] PANDYA D S,ZAVERI M A. Frame based approach for automatic event boundary detection of soccer video using optical flow [C]//Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications. Kuching, Malaysia: IEEE,2017:402-406.

[12] FAKHAR B,KANAN H R,BEHRAD A. Learning an event-oriented and discriminative dictionary based on an adaptive label-consistent K-SVD method for event detection in soccer videos [J]. Journal of Visual Communication and Image Representation,2018,55:489-503.

[13] MORRA L,MANIGRASSO F,CANTO G,et al. Slicing and slicing soccer:automatic detection of complex events from spatiotemporal data[C]//Proceedings of the International Conference on Image Analysis and Recognition. Cham: Springer, 2020: 107-121.

[14] GAO X,LIU X,YANG T,et al. Automatic Key Moment Extraction and Highlights Generation Based on Comprehensive Soccer Video Understanding[C]//Proceedings of the IEEE International Conference on Multimedia & Expo Workshops. London,UK:IEEE,2020:1-6.

[15] TOMEI M,BARALDI L,CALDERARA S,et al. RMS-Net:Regression and Masking for Soccer Event Spotting[C]//Proceedings of the 25th International Conference on Pattern Recognition. Milan,Italy:IEEE,2021:7699-7706.

[16] CIOPPA A, DELIÈGE A, GIANCOLA S, et al. A Context-Aware Loss Function for Action Spotting in Soccer Videos [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 13123-13133.

[17] JIANG H, LU Y, XUE J. Automatic Soccer Video Event Detection Based on a Deep Neural Network Combined CNN and RNN [C]// Proceedings of the 28th International Conference on Tools with Artificial Intelligence. San Jose, CA, USA: IEEE, 2016: 490-494.

[18] SONG C, RASMUSSEN C. Multi-camera Temporal Grouping for Play/Break Event Detection in Soccer Games [C]// Proceedings of the International Symposium on Visual Computing. Cham: Springer, 2019: 231-243.

[19] FENG N, SONG Z, YU J, et al. SSET: a dataset for shot segmentation, event detection, player tracking in soccer videos [J]. Multimedia Tools and Applications, 2020, 79(39): 28971-28992.

[20] WANG T, FENG N, YU J, et al. Shot Boundary Detection Through Multi-stage Deep Convolution Neural Network [C]// Proceedings of the International Conference on Multimedia Modeling. Cham: Springer, 2019: 456-468.

[21] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv: 1409. 1556, 2014.

[22] TRAN D, BOURDEV L D, FERGUS R, et al. Learning Spatio-temporal Features with 3D Convolutional Networks [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2015: 4489-4497.

[23] WU L, YU J, HE Y, et al. Soccer video shot classification and playfield zone detection [J]. Journal of Image and Graphics, 2017, 22(2): 265-272.

[24] CHENG P, FENG J, MA H J, et al. Soccer Video Scene Classification Algorithms Based on C3D [J]. Computer System Application, 2019, 28(12): 158-164.

[25] SARKAR S, ALI S, CHAKRABARTI A. Shot Classification and Replay Detection in Broadcast Soccer Video [C]// Proceedings of the Advanced Computing and Systems for Security. Singapore: Springer, 2020: 57-66.

[26] YU J, LEI A, HU Y. Soccer Video Event Detection Based on Deep Learning [C]// Proceedings of the International Conference on Multimedia Modeling. Cham: Springer, 2019: 377-389.

[27] VANDERPLAETSE B, DUPONT S. Improved Soccer Action Spotting using both Audio and Video Streams [C]// Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, WA, USA: IEEE, 2020: 3921-3931.



HUA Xiaofeng, born in 1996, postgraduate. Her main research interest is soccer video event detection.



YU Junqing, born in 1975, Ph.D, professor, Ph.D supervisor. His main research interests include content-based video analysis, indexing and retrieval, multi-core computing and stream compilation, video emotion computing, network security and big data processing.

(责任编辑: 何杨)