

中国图象图形学学会

)

北京应用物理与计算数学研究所

中国图象 图形学报



中国多媒体大会(ChinaMM2020)专刊

Ş





中国精品科技期刊 中国国际影响力优秀学术期刊 中国科技核心期刊 中文核心期刊

版权声明

凡向《中国图象图形学报》投稿,均视 为同意在本刊网站及CNKI等全文数据 库出版,所刊载论文已获得著作权人的 授权。本刊所有图片均为非商业目的使 用,所有内容,未经许可,不得转载或

Copyright

All rights reserved by Journal of of Remote Sensing and Digital (including but not limited text,

中国图象图形学报

刊名题字: 宋健 月刊(1996年创刊)

主管单位 中国科学院 主办单位 中国科学院空天信息创新研究院 中国图象图形学学会

北京应用物理与计算数学研究所

主 编	吴一戎
编辑出版	《中国图象图形学报》编辑出版委员会
通信地址	北京市海淀区北四环西路19号
邮编	100190
电子信箱	jig@aircas.ac.cn
电 话	010-58887035
网址	www.cjig.cn
<u>+++++</u>	
) 古友伯(夏165 宋期 冏/ 安子201/02185

总发行	, 北京报刊发行局
订 则	了全国各地邮局
海外发行	,中国国际图书贸易集团有限公司
	(邮政信箱:北京399信箱 邮编:100048)
印刷装订	「 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences Sponsored by Aerospace Information Research Institute, CAS China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China Zip code 100190 E-mail jig@aircas.ac.cn Telephone 010-58887035 Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals Domestic All Local Post Offices in China **Overseas** China International Book Trading Corporation (P.O.Box 399, Beijing 100048, P.R.China)) Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB ISSN 1006-8961 CODEN ZTTXFZ

国外发行代号 M1406 国内邮发代号 82-831 国内定价 60.00元



综述



多模态多层次事件网络的诺言检测	
李莎,张怀文,钱胜胜,方全,徐常胜 ······ 164	8
多模态零样本人体动作识别	
吕露露,黄毅,高君宇,杨小汕,徐常胜 ······165	8
足球视频球员感知跟踪算法	
冯思佳,宋子恺,于俊清,何云峰,管涛	8
融合时空图卷积的多人交互行为识别	
成科扬,吴金霞,王文杉,荣兰,詹永照 ······168	1
时序特征融合的视频实例分割	
黄泽涛,刘洋,于成龙,张加佳,王轩,漆舒汉169	2

医学图像处理

多尺度深度特征提取的肝脏肿瘤CT图像分类	
毛静怡,宋余庆,刘哲 ·······17	04
3D多尺度深度卷积神经网络肺结节检测	
孙华聪,彭延军,郭燕飞,张晓庆 ······ 17	16
结合多通道注意力的糖尿病性视网膜病变分级	
顾婷菲,郝鹏翼,白琮,柳宁17	26



类别语义相似性监督的小样 本图像识别(第1594页)



像素聚合和特征增强的任意 形状场景文本检测(第1614 页)



时序特征融合的视频实例分 割(第1692页)

Volume 26, Number 7 Published July 16, 2021 CONTENTS JOURNAL OF IMAGE AND GRAPHICS



Few shot image recognition based on class semantic similarity supervision(P1594)



Arbitrary shape scene-text detection based on pixel aggregation and feature enhancement (P1614)



Video instance segmentation based on temporal feature fusion (P1692)

Review

Survey on compressive sensing video stream for uplink streaming media					
Liu Hao, Huang Rong, Yuan Haodong	1545				

Image Processing and Coding

Structure-preserving hashing with coupled projections for cross-modal retrieval				
Min Kangling, Zhang Guobin, Wang Lei, Li Danping	1558			
Locality-sensitive hashing approach based on semantic space for visual retrieval				
Huang Xiaoyan, Sun Bin, Yang Zhanyuan, Zhu Yingying, Tian Qi	1568			

Image Analysis and Recognition

Visual relationship detection-based emergency early-warning description generation in electric
power industry
Gao Ming, Zuo Hongqun, Bai Fan, Tian Qingyang, Ge Zhifeng, Dong Xingning, Gan Tian … 1583
Few shot image recognition based on class semantic similarity supervision
Xu Pengbang, Sang Jitao, Lu Dongyuan
Object distance estimation based on stereo regional disparity regression
Zhang Yufeng, Li Yuxi, Zhao Mingbi, Yu Xiaoyuan, Zhan Yunlong, Lin Weiyao
Arbitrary shape scene-text detection based on pixel aggregation and feature enhancement
Shi Guangchen, Wu Yirui ····· 1614
Phase consistency guided full-reference panoramic image quality assessment algorithm
Xia Yumeng, Wang Yongfang, Wang Chuang ····· 1625

Image Understanding and Computer Vision

Temporal action detection based on feature pyramid hierarchies
He Jiayu, Lei Jun, Li Guohui 1637
Multi-modal multi-level event network for rumor detection
Li Sha, Zhang Huaiwen, Qian Shengsheng, Fang Quan, Xu Changsheng 1648
Multimodal-based zero-shot human action recognition
Lyu Lulu, Huang Yi, Gao Junyu, Yang Xiaoshan, Xu Changsheng
Players-aware tracking algorithm in soccer video
Feng Sijia, Song Zikai, Yu Junqing, He Yunfeng, Guan Tao
Multi-person interaction action recognition based on spatio-temporal graph convolution
Cheng Keyang, Wu Jinxia, Wang Wenshan, Rong Lan, Zhan Yongzhao
Video instance segmentation based on temporal feature fusion
Huang Zetao, Liu Yang, Yu Chenglong, Zhang Jiajia, Wang Xuan, Qi Shuhan

Medical Image Processing

CT image classification of liver tumors based on multi-scale and deep feature extraction	
Mao Jingyi, Song Yuqing, Liu Zhe ····· 17	'04
3D multi-scale deep convolutional neural networks in pulmonary nodule detection	
Sun Huacong, Peng Yanjun, Guo Yanfei, Zhang Xiaoqing	'16
Diabetic retinopathy grading based on multi-channel attention	
Gu Tingfei, Hao Pengyi, Bai Cong, Liu Ning	26

中图法分类号:TP391 文献标识码: A 文章编号: 1006-8961(2021)07-1668-13

论文引用格式: Feng S J, Song Z K, Yu J Q, He Y F and Guan T. 2021. Players-aware tracking algorithm in soccer video. Journal of Image and Graphics,26(07):1668-1680(冯思佳,宋子恺,于俊清,何云峰,管涛. 2021. 足球视频球员感知跟踪算法. 中国图象图形学报,26(07):1668-1680) [DOI:10.11834/jig.200507]

足球视频球员感知跟踪算法

冯思佳¹,宋子恺¹,于俊清^{1,2*},何云峰¹,管涛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

摘 要:目的 足球比赛视频中的球员跟踪算法为足球赛事分析提供基础的数据支持。但足球比赛中球员跟踪存 在极大的挑战:球员进攻、防守和争夺球权时,目标球员可能产生快速移动、严重遮挡和周围出现若干名干扰球员 的情况,目前仍没有一种能够完美解决足球比赛中球员跟踪问题的算法。因此如何解决足球场景中的困难,提升 球员跟踪的准确度,成为当前研究的热点问题。方法 本文在分析足球比赛视频中球员目标特点的基础上,通过 融合干扰项感知的颜色模型和目标感知的深度模型,提出并设计了一种球员感知的跟踪算法。干扰项感知的颜色 模型分别提取目标、背景和干扰项的颜色直方图,利用贝叶斯公式得到搜索区域中每个像素点属于目标的似然概 率。目标感知的深度模型利用孪生网络计算搜索区域与目标的相似度。针对跟踪漂移问题,使用全局跟踪器和局 部跟踪器分别跟踪目标整体和目标上半身,并且在两个跟踪器的跟踪结果出现较大差异的时候分析跟踪器有效性 并进行定位修正。结果 在公共的足球数据集上将本文算法与10个其他跟踪算法进行对比实验,同时对于文本 算法进行了局部跟踪器的消融实验。实验结果表明,球员感知跟踪算法的平均有效重叠率达到了0.5603,在存在 同队球员和异队球员干扰的情况下,本文算法比排名第2的算法的有效重叠率分别高出3.7%和6.6%,明显优于 其他算法,但是由于引入了干扰项感知的颜色模型、目标感知的深度模型以及局部跟踪器等模块增加了算法的时 间复杂度,导致本文算法跟踪速度较慢。结论 本文总结了跟踪算法的整体流程并分析了实验结果,认为干扰项 感知、目标感知和局部跟踪这3个策略在足球场景中的球员跟踪问题中起到了重要的作用,为未来在足球球员跟 踪领域研究的继续深入提供了参考依据。

关键词:计算机视觉;图像处理;目标跟踪;足球球员跟踪;干扰项感知;目标感知;局部跟踪

Players-aware tracking algorithm in soccer video

Feng Sijia¹, Song Zikai¹, Yu Junqing^{1,2*}, He Yunfeng¹, Guan Tao¹

School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China;
 Center of Network and Computation, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract: **Objective** Target object tracking is important in computer vision. Player-tracking algorithms in broadcast soccer videos provide basic data support for the analysis of soccer matches. Several challenges occur in soccer player tracking, including a rapid move of the target player, occlusion, and disturbance of similar players when they attack, defend, and scramble for the ball. However, no perfect tracking algorithm specifically for soccer video is available. The following challenges remain in the player tracking of broadcast soccer videos: 1) A small patch of target players in the video frame is not conducive to feature extraction. 2) Similar players often interfere with the target player. 3) Occlusion of the target player

收稿日期:2020-08-24;修回日期:2020-12-30;预印本日期:2021-01-06

^{*}通信作者:于俊清 yjqing@hust.edu.cn

基金项目:国家自然科学基金项目(61572211)

Supported by: National Natural Science Foundation of China (61572211)

by other players often occurs, requiring the algorithm to distinguish intra-class targets. 4) Relocating the target after tracking drift is difficult. Thus, a prevalent topic in current research is how to handle the challenges in the soccer scene and improve the accuracy of player tracking. Method Based on a depth analysis of the characteristics of a soccer player, we propose and design a player-aware tracking algorithm by fusing a distractor-aware color model and the target-aware deep model. In the color model, the color histogram of the target player, background, and distractors are extracted. The color model based on the Bayesian classifier aims to identify the foreground target from the background by color information in the search region. Three primary color components in the RGB color space are divided into 16 color regions by uniform quantization. The color histogram of the corresponding region can be obtained by calculating the number of pixels in each color interval. Distractors are non-target candidate regions whose similarity scores are larger than a certain threshold in the response map. As with the foreground-background color model, the color histogram of the target and distractor is counted, and the likelihood probability that the pixel belongs to the target in the target-distractor item is obtained. In the deep model, Siamese networks are adopted to calculate the similarity between the search and target regions. The target-aware deep model embeds deep features into the Siamese network, calculates the similarity between the output of the template branch and detects branches to obtain a response map of the search region. The well-known Visual Geometry Group (VGG) feature extraction network is adopted as a backbone network. In feature space, each channel of feature represents a different feature-representation capability, and specific combinations of features can recognize specific categories. The response of one category only focuses on specific deep-feature channels but not all feature channels. For the current tracking player, we design a small regression network to select feature channels related to the tracking player from VGG deep features. The structure of the small regression network is composed of one convolution layer with one convolution kernel. The size of the convolution kernel is the same as that of the target feature. The regression network aims to fit the features of the target sample to Gaussian distribution. In addition, to solve the problem of tracking drift, a global-local tracking strategy is designed to track the entire target and upper part of the target. Both global and local trackers have the same network architecture, including a distractor-aware color model branch and target-aware deep model branch. When a great difference in tracking results exists between the global and local trackers, the effectiveness of each tracker is analyzed and location revision is performed. In online tracking, both global and local trackers are used to track the whole and upper part of the target. When one tracker drifts, another is used to revise the target position. According to the intersection over union of the target of the global and local trackers, the tracking results can be classified into stable and unstable states. A stable state is when the intersection over union of the target boxes of the local and global trackers is greater than a certain threshold, while an unstable state indicates less than that threshold. In the unstable state, the following factors are considered simultaneously to analyze the tracker: main color similarity of the target in the current and initial frames, maximum response value of the response map, and moving distance from the center of the previous frame to the current frame. The lower the main color similarity, the more likely the tracker will be lost to the non-target player. The smaller the maximum response value of the response map, the lower is the reliability of the tracker. The moving distance of the tracker box is greater than a certain threshold, which indicates that the tracker is likely to have a sudden tracking drift in the current frame. Result We select 10 state-of-the-art tracking algorithms and compare them with the proposed algorithm on the public soccer dataset. The ablation experiment on the global-local tracking strategy is expanded. Experimental results show that the average valid overlap rate of the proposed tracking algorithm is 0.560 3, and when the target player is occluded by players in the same team and different teams, the average valid overlap rate of the proposed algorithm is 3.7% and 6.6% higher than that of the second-ranked algorithm, respectively. The evaluation results demonstrate that the player-aware tracking algorithm is more effective than other algorithms in addressing the disturbance by other similar players. However, the tracking speed is slow due to the increase of computational complexity by introducing the color model, deep model, and global-local tracking strategy. Conclusion We summarize the entire process of the proposed tracking algorithm and analyze the experimental results. Three strategies, namely, distractor-aware color model, target-aware deep model, and global-local tracking strategy, are demonstrated to play a crucial role in player tracking. In terms of the color model, the color histogram of the target player, background, and distractor are extracted, and the likelihood probability that each pixel in the search region belongs to the target is calculated by using the Bayesian formula. In terms of the deep model, a small regression network is adopted to select feature channels related to the target object from the deep feature, and the Siamese network is used to calculate the similarity between the search region and target object. To alleviate tracking drift, we use the global-local strategy to track the whole target and upper body of the target so that the failure location can be revised. This study provides a basic reference for further research on player tracking in broadcast soccer videos.

Key words: computer vision; image processing; object tracking; player tracking; distractor aware; target aware; globallocal tracking strategy

0 引 言

体育视频中的运动目标跟踪技术对于智能体育 视频系统的发展与应用具有重要意义。作为拥有极 高关注度的体育比赛之一,足球比赛视频中的球员 跟踪算法吸引了不少研究人员的目光。足球比赛视 频中的球员跟踪是指在初始帧给出目标球员的位置 信息后,利用跟踪算法计算后续帧中目标球员的位置 信息的技术。足球比赛视频中的球员跟踪为赛事 分析提供了基础的数据支持:利用球员的移动距离、 奔跑速度等参数帮助教练进行比赛数据分析;利用 球员的运动轨迹、瞬间速度等参数帮助裁判对有争 议的判罚进行裁定。

从 20 世纪 50 年代初起,目标跟踪取得丰富的 研究成果,但是至今仍没有出现一种能够胜任所有 场景的跟踪任务的算法。足球比赛视频中的球员跟 踪依然存在以下挑战:1)目标球员所占面积较小, 不利于目标球员的特征提取;2)目标球员周围常有 相似球员干扰;3)目标球员可能被其他球员遮挡; 4)跟踪漂移后无法重定位。

针对足球比赛视频这一特定跟踪场景,结合足 球领域特点,提出并设计了一种球员感知的跟踪算 法,利用手工特征和深度特征相结合的方式来增强 球员的特征表达能力以缓解球员面积小且特征不明 显的问题,利用干扰项感知的颜色模型来排除目标 球员周围相似球员的干扰问题,采用局部跟踪器以 及定位修正来解决球员被遮挡时可能引起的跟踪漂 移问题,本文算法在公开的足球数据集上取得了优 异的成绩。主要贡献如下:

 1)为了充分利用手工特征和深度特征的互补 性,分别提取颜色特征和目标感知的深度特征描述 目标球员。

2)通过融合干扰项感知的颜色模型和目标感知的深度模型,提出并设计了一种球员感知的跟踪

算法。

3)使用全局和局部跟踪器分别对目标整体和 目标上半身独立跟踪,在跟踪漂移时进行定位修正。

1 相关工作

通用的目标跟踪算法层出不穷。以核化相关滤 波器(kernelized correlation filters, KCF)(Henriques 等,2015)为代表的基于相关滤波的跟踪算法在实 时性方面表现优秀,但不足以应对复杂背景干扰、相 似物遮挡等情况。以深度学习跟踪器(deep learning tracker, DLT)(Wang 和 Yeung,2013)为代表的基于 深度学习的跟踪算法则更侧重于性能,大多无法满 足实时性。以全卷积孪生网络(fully-convolutional siamese networks, SiamFC)(Bertinetto 等,2016b)为 代表的使用孪生神经网络的跟踪算法兼顾了性能和 实时性。

与通用的目标跟踪算法不同,足球视频中场景 的复杂性和目标运动状态的不确定性限制了球员跟 踪算法的研究。足球比赛视频中的跟踪算法大多结 合检测算法,并且与足球比赛视频中的目标特点紧 密联系。

Heydari和 Moghadam (2012)提取 YCbCr 颜色 空间下的特征值和区域面积作为特征,在第1阶段 利用 K 均值聚类算法获取足球场地主色,结合足球 领域知识和形态学操作去除场地上的非球员区域; 在第2阶段利用多层感知神经网络将球员区域进一 步细分为目标球员区域和非目标球员区域。娄娜 等人(2007)利用颜色统计信息和像素边缘特征对 球员区域进行检测,并为球员区域设置编号、球员模 板坐标、区域中球员的编号和区域内球员个数4个 属性,结合图像匹配进行球员跟踪。付龙(2015)利 用卡尔曼滤波与局部区域匹配相结合的方法进行多 球员跟踪,根据相邻帧之间的球员质心和球员面积 的改变,判断当前帧是否发生遮挡,并在遮挡结束后

及时利用匹配策略对球员进行重定位。王勋 (2017)在 HSV 颜色空间下去除球场主色,提取上下 两块球员主色直方图和哈尔特征,利用改进的传统 在线多示例学习方法对球员进行跟踪。Lu 等人 (2011)利用可变形局部模型(deformable parts model, DPM)对足球场上的人物进行分类,建立条件随 机场(conditional random fields, CRF)模型进行联合 概率推理。Morais 等人(2012)利用多个提前放置 在足球场地上的固定摄像机拍摄到的画面,融合多 摄像机的参数生成外观模型,投射到球场平面上以 生成多峰双向概率函数,该球场平面概率能够反映 球场平面上球员可能出现的位置。Seo 等人(1997) 在去除足球主场地和分割出球员区域的基础上,利 用卡尔曼滤波与模板匹配相结合的方法对球员进行 跟踪,并通过直方图反向投影的方式处理遮挡情况。 马月洁等人(2018)利用目标检测数据集和足球视 频数据集对全卷积孪生网络进行离线训练,获取适 合球员跟踪任务的深度特征。Liu 等人(2013)认为 球员的运动与其附近的球员有高度相关性,利用层 级间联系在长时间内跟踪多个球员。Li 和 Flierl (2012)提出一种基于 SIFT (scale-invariant feature transform)特征多视角的方式进行多球员跟踪,假设 未来的比赛视频会有多个视角的高分辨率摄像机拍 摄,能提供自由视角的网上观看体验,球员精确的 3D 信息被高效地用于跟踪,通过共享所有摄像机的 3D 信息解决遮挡问题。Baysal 和 Duygulu(2016)提 出了模型场地粒子的概念,把足球场地用一组密集 采样的粒子集合来表示,使球员位置离散化,从全局 上联合目标外观和运动模型进行位置可能性计算。 Kataoka和 Aoki(2011)在分割出足球场地的基础 上,利用粒子滤波跟踪球员,利用 Real AdaBoost分 类器解决球员之间相互遮挡的问题,结合摄像机运 动获得球员在球场上的运动轨迹。Mazzeo等人 (2008)在利用背景去除法分割出运动球员和利用 无监督聚类算法对运动球员分类的基础上,采用最 大后验概率估计的方法跟踪球员。Najafzadeh等人 (2015)提出将球员平均分为大小相同的头颈、躯 干、脚3个块,分别统计每个块的颜色直方图,利用 Bhattacharyya距离计算每个块的权重,以实现适应 块的模型更新。Bastanfard等人(2019)在 YCbCr 颜 色空间下检测并去除阴影,利用显著图算法检测球 场区域,利用边缘检测、颜色检测和最大类间方差三 大算法检测场上球员,结合改进的粒子滤波算法进 行球员跟踪。

现有的足球球员跟踪算法大多在通用跟踪算法 的基础上采用球场的颜色特征以增强跟踪器对足球 场景的适应性,但是足球场景中球员面积小、特征不 明显以及比赛中经常出现的球员遮挡情况没有得到 很好的解决。本文针对这些挑战设计了球员感知的 跟踪算法,通过引入目标感知的深度模型、干扰项感 知的颜色模型以及局部跟踪器的方法提高了球员跟 踪的准确率。

2 球员感知跟踪算法

如图 1 所示,通过融合干扰项感知的颜色模型 和目标感知的深度模型,提出并设计了一种球员感 知的跟踪算法(players aware yracking model, PAT)。



Fig. 1 The flow chart of proposed players aware tracking algorithm

中国图象图形学报 JOURNAL OF IMAGE AND GRAPHICS

目标感知中的目标指目标球员,干扰项感知中的 干扰项指目标球员周围的非目标球员。干扰项感知 指在颜色模型中不仅考虑目标球员和背景的颜色差 异,还考虑目标球员和干扰项的颜色差异。目标感知 指在深度模型中使用与当前跟踪球员相关的深度特 征,而非普通的 VGG-16(Visual Geometry Group-16) (Simonyan 和 Zisserman,2015)深度特征。针对跟踪 漂移问题,在线跟踪过程时使用全局跟踪器和局部 跟踪器分别用于跟踪目标整体和目标上半身,当两 个跟踪器的跟踪结果之间的距离大于一定阈值时, 进行定位修正。

2.1 干扰项感知的颜色模型

利用目标颜色信息的跟踪算法,极容易受到与 目标颜色相似的非目标物体的干扰。足球比赛中, 目标球员周围常有颜色属性非常相似的同队球员出现,在目标周围出现的同队球员极有可能成为干扰项。如图2所示,干扰项感知的颜色模型利用贝叶 斯分类器对目标—背景以及目标—干扰项建模。





2.1.1 目标—背景颜色模型

基于贝叶斯分类器的颜色模型首先对目标—背 景建模,该模型直接在搜索图像中统计颜色信息:采 取均匀量化的方式将 RGB 颜色空间中的三原色分 量各划分成16 个颜色区间,计算落在每个颜色区间 的像素数量即可得到对应图像的颜色直方图。

记统计区域为S,目标球员区域为T,背景区域为 B,当前位置像素为x,当前位置像素所属颜色区间 b_x ,计算搜索图像中像素x属于目标的似然概率为

$$P(x \in T | T, B, b_x) \approx$$

$$\frac{P(b_x | x \in T) P(x \in T)}{P(b_x | x \in T) P(x \in T) + P(b_x | x \in B) P(x \in B)}$$
(1)

像素点属于目标或者背景的先验概率可以利用 颜色直方图简化表达为

$$P(\boldsymbol{b}_{x} \mid x \in \boldsymbol{T}) \approx \frac{H_{T}^{s}(\boldsymbol{b}_{x})}{|\boldsymbol{T}|}$$
(2)

$$P(\boldsymbol{b}_{x} \mid x \in \boldsymbol{B}) \approx \frac{H_{B}^{S}(\boldsymbol{b}_{x})}{|\boldsymbol{B}|}$$
(3)

式中, $H_T^s(\boldsymbol{b}_x)$ 表示在目标颜色直方图中 \boldsymbol{b}_x 颜色区 间所对应的统计值, $H_B^s(\boldsymbol{b}_x)$ 表示在背景颜色直方 图中 \boldsymbol{b}_x 颜色区间所对应的统计值, $|\boldsymbol{T}|$ 表示目标面 积, $|\boldsymbol{B}|$ 表示背景面积。 $P(x \in \boldsymbol{T})$ 和 $P(x \in \boldsymbol{B})$ 可 以用面积占比来表示,即

$$P(x \in \mathbf{T}) \approx \frac{|\mathbf{T}|}{|\mathbf{T}| + |\mathbf{B}|}$$
(4)

$$P(x \in \boldsymbol{B}) \approx \frac{|\boldsymbol{B}|}{|\boldsymbol{T}| + |\boldsymbol{B}|}$$
(5)

结合式(2)—(5),式(1)可以简化表达为 $P(x \in T | T, B, b_x) =$

$$\begin{cases} \frac{H_T^S(\boldsymbol{b}_x)}{H_T^S(\boldsymbol{b}_x) + H_B^S(\boldsymbol{b}_x)} & \boldsymbol{b}_x \in \{\boldsymbol{b}(\boldsymbol{T} \cup \boldsymbol{B})\} \\ 0.5 & \ddagger \& \end{cases}$$
(6)

当 $b_x \notin \{b(T \cup B)\}$ 时,该颜色区间在目标区域和背景区域中都不曾出现,颜色模型无法从先验概率中分辨该颜色更倾向于哪一方,故该情况下像素x属于目标的似然概率为0.5。

2.1.2 目标—干扰项颜色模型

干扰项生成工作将在在线跟踪中详细说明,现 假设已知干扰项集合。与目标—背景颜色模型一 样,分别统计目标和干扰项的颜色直方图,在目标— 干扰项模型中像素 x 属于目标的似然概率为

$$P(x \in T \mid T, D, b_{x}) = \begin{cases} \frac{nH_{T}^{s}(b_{x})}{nH_{T}^{s}(b_{x}) + H_{D}^{s}(b_{x})} & b_{x} \in \{b(T \cup D)\} \\ 0.5 & \text{ 其他} \end{cases}$$
(7)

式中, D 代表所有的干扰项, n 代表干扰项个数。 $H_D^s(\boldsymbol{b}_x)$ 代表在干扰项颜色直方图中 \boldsymbol{b}_x 颜色区间所 对应的统计值。

在得到目标—背景颜色模型和目标—干扰项颜 色模型后,结合目标—背景模型和目标—干扰项模 型,得到干扰项感知的颜色模型,即

$$P(x \in \boldsymbol{T} \mid \boldsymbol{b}_{x}) =$$

$$\lambda_{1}P(x \in \boldsymbol{T} \mid \boldsymbol{T}, \boldsymbol{B}, \boldsymbol{b}_{x}) +$$

$$(1 - \lambda_{1})P(x \in \boldsymbol{T} \mid \boldsymbol{T}, \boldsymbol{D}, \boldsymbol{b}_{x})$$
(8)

通过干扰项感知的颜色模型,计算搜索图像内 各像素点属于目标的似然概率,得到搜索图像对应 的像素级概率图,通过高斯投票机制进一步将像素 级概率图转化为响应图。响应图中各响应值对应着 一个固定尺寸的目标候选框。在计算目标候选框的 响应得分时,不仅考虑了像素颜色属于目标的似然 概率,还考虑了像素和上一帧目标中心点的距离。 计算搜索图像的响应图*M*₁,即

$$\boldsymbol{M}_{1} = (\boldsymbol{s}_{v}(\boldsymbol{T}_{i,i}) \ \boldsymbol{s}_{d}(\boldsymbol{T}_{i,i}))$$
(9)

$$s_v(\boldsymbol{T}_{t,i}) = \sum_{x \in \boldsymbol{T}_{t,i}} P_t(x \in \boldsymbol{T} \mid \boldsymbol{b}_x) \quad (10)$$

$$s_d(\mathbf{T}_{t,i}) = \sum_{x \in \mathbf{T}_{t,i}} e^{-\frac{||x-c_{t-1}||^2}{2\sigma^2}}$$
(11)

式中, $T_{i,i}$ 代表第 t 帧的第 i 个候选框, $s_{i}(\cdot)$ 代表候

选框的投票得分, $s_{d}(\cdot)$ 代表候选框的距离得分。 c_{t-1} 代表第t-1帧的目标中心位置, σ 为高斯函数 中的标准方差参数。

2.2 目标感知的深度模型

如图 3 所示,目标感知的深度模型将目标感知 的深度特征嵌入孪生网络,将模板分支(template) 输出的特征通过由单层卷积网络(Conv)和 L2 损失 函数(L2 Loss)组成的回归网络生成目标激活向量 (activate vector),通过目标激活向量作用于原始深 度特征,便可以得到目标感知的深度特征,目标感知 的深度特征作为卷积核,与检测分支(search region) 输出的目标感知的深度特征做相关操作(Corr),得 到搜索图像的响应图 *M*₂。



Fig. 3 Target aware deep model

2.2.1 目标感知的深度特征

VGG-16(Simonyan 和 Zisserman, 2015)在 ImageNet 上训练得到的深度神经网络模型学习到一组 丰富的、有辨别力的特征来识别不同类别的物体。 VGG-16 共包含 13 个卷积层,4 224 个卷积滤波器,每 个卷积滤波器捕获一个特定的特征模式。VGG-16 中所有的卷积滤波器构造了一个包含 1 000 个类别 的特征空间,在该特征空间中,深度特征的每一个维 度代表了不同的特征组合,特定的特征组合可以识 别特定的类别。一个特定的类别的分类响应只集中 在特定的深度特征通道上,而不是所有深度特征 通道。

针对当前跟踪球员,利用小型回归网络在 VGG-16 深度特征中选择与该跟踪球员相关的特征通道。小型回归网络结构试图将目标样本特征通过以下方式 拟合高斯分布: 1)输入初始帧搜索图像的深度特征作为目标 样本。

2)经过一个卷积层,卷积核大小为目标图像特征的大小,卷积核个数为1。

3) 真实响应值为高斯标签值, 越靠近目标中心 的位置标签值越高, 越远离目标中心的位置标签值 越低。输出损失函数为 L2 损失, 损失函数为

 $L = \| Y(i,j) - W * X(i,j) \|^2 + \lambda \| W \|^2$ (12) 式中, $Y(i,j) = e^{-\frac{2+j^2}{2\sigma^2}}$, (*i*,*j*) 代表该像素点到目标 中心的偏移量, Y(i,j) 代表该像素点的高斯响应 值, W 为卷积参数, X(i,j) 为目标图像的 VGG-16 深度特征, λ 为正则化参数。

等模型收敛后,对目标图像的 VGG-16 深度特 征反向求导,获得深度特征各个通道的梯度值,各通 道的梯度值代表了该特征通道对于拟合高斯分布任 务的贡献,即 中国图象图形学报 JOURNAL OF IMAGE AND GRAPHICS

$$\frac{\partial L}{\partial X} = \sum_{i,j} 2(Y(i,j) - W * X(i,j)) \times W$$
(13)

各个特征通道的重要性度量方法为

$$\theta_k = G_{AP} \left(\frac{\partial L}{\partial X_k} \right) \tag{14}$$

式中, k代表特征通道, θ_k 代表第k个特征通道的重要程度, $G_{AP}(\cdot)$ 为全局平均池化函数。

根据特征通道重要性,生成目标激活向量v,该 向量维度和样本特征通道数保持一致。目标激活向 量v 仅激活排在前 m 位的特征通道,其余通道不激 活,即

$$v_{k} = \begin{cases} 1 & rank(\theta_{k}, \theta_{1}; \theta_{m}) \leq m \\ 0 & \pm \ell h \end{cases}$$
(15)

计算目标感知的深度特征

$$X_{\text{target activate}} = X_{\text{original}} \cdot \nu$$
 (16)

式中, X_{original} 表示激活前的原始特征,即从 VGG-16 直接提取出的特征, $X_{\text{target_activate}}$ 表示激活后的特征。

2.2.2 孪生网络

孪生网络将目标跟踪问题转化为相似性学习问题,两路分支分别为模板分支和检测分支,两分支共 享一个神经网络模型和一套模型参数。

对于模板分支,提取初始帧目标图像的 VGG-16 第 10 层深度特征,将该深度特征与目标激活向量 v 相乘,得到初始帧目标图像的目标感知的深度特征。 对于检测分支,同样地,提取当前帧搜索图像的 VGG-16 第 10 层深度特征,将该深度特征与目标激 活向量 v 相乘,得到当前帧搜索图像的目标感知的 深度特征。

将模板分支输出的目标感知的深度特征作为卷 积核,与检测分支输出的目标感知的深度特征做卷 积操作,得到搜索图像的响应图 *M*₂,即

$$\boldsymbol{M}_{2} = (\boldsymbol{X}_{1} \cdot \boldsymbol{v}) \ast (\boldsymbol{X}_{\iota} \cdot \boldsymbol{v})$$
(17)

卷积操作相当于对搜索图像进行密集的滑动窗 口评估,每一次评估计算了滑动窗口图像特征与模 板图像特征的相关性,如果这两个图像所描绘的对 象高度相似则返回高分,否则返回低分。

目标跟踪任务需要在搜索图像中进行像素级别的定位,而卷积操作得到的搜索区域的响应图 *M*₂ 比搜索区域本身小很多,因此需要对响应图进行上 采样。上采样的方式有多种,如双线性插值、反卷积 和反池化等,故采用双线性插值将响应图重新映射 回搜索图像。

2.3 融合模型

融合干扰项感知的颜色模型和目标感知的深度 模型,即可得到 PAT。在线跟踪时,同时使用全局跟 踪器和局部跟踪器跟踪目标整体和目标上半身,在 一个跟踪器发生跟踪漂移时,使用另一个跟踪器进 行定位修正。

2.3.1 初始化

全局跟踪器的初始目标球员位置用一个矩形目标框表示为 (x_g, y_g, w_g, h_g) ,局部跟踪器的初始局部目标位置则用矩形目标框表示为 $(x_g, y_g, w_g, h_g/2)$ 。

对于全局跟踪器,初始化一个干扰项感知的深 度模型、一个目标感知的深度模型和一个目标主色。 根据颜色直方图中颜色块索引相对应的颜色出现的 频率高低对颜色块索引进行排序,得到4096×1的 向量,提取其中前 n 个颜色索引构成一个 n×1的向 量,作为目标主色。目标主色用于辅助分析跟踪器 有效性。

对于局部跟踪器的初始化工作和全局跟踪器完 全一致。

2.3.2 在线跟踪

对于局部跟踪器,为了对目标球员进行简单尺 度估计,利用双线性插值得到3个不同尺度的搜索 图像。3个不同尺度的搜索图像经过干扰项感知的 颜色模型得到多尺度颜色模型响应图*M*₁;经过目标 感知的深度模型得到多尺度深度模型响应图*M*₂。 将相同尺度因子所对应的颜色模型响应图和深度模 型响应图加权融合,得到多尺度融合响应图,即

 $\boldsymbol{M} = \boldsymbol{\lambda}_2 \, \boldsymbol{M}_1 + (1 - \boldsymbol{\lambda}_2) \, \boldsymbol{M}_2 \tag{18}$

在多尺度融合响应图中,最大响应值所对应的 尺度因子为当前帧目标的尺度;最大响应值对应的 位置为当前帧目标的中心位置,两者结合得到当前 帧的目标框。

由于同一个镜头中的非目标球员和目标球员的 尺寸相差不大,故干扰项尺寸和当前目标尺寸保持 一致。在最大响应值所在的融合响应图中,干扰项 生成过程步骤为:

 1)根据融合响应图 M,选择主峰位置对应候选 框为当前目标框,记最大置信度为 C_{max},并将其加 入干扰项候选集合 P中。

2)选择融合响应图中置信度大于β・C_{max}的除 当前目标框以外的候选框,将其加入干扰项候选集 合 P, 根据置信度对 P 中候选框进行降序排序。

3)选择当前 P 中最大置信度对应的候选框 d, 将其从 P 中删除并加入干扰项集合 D,计算 d 与 P中所有候选框 d_i 的交并比(intersection over union, IoU),若交并比大于阈值则将该 d_i 从 P 中删除。

4) 重复步骤 3), 直到当前 P 为空。

5)将当前 D 中置信度最大的侯选框删除。

局部跟踪器的在线跟踪算法和全局跟踪器完全 一致。

2.3.3 定位修正

根据全局跟踪器和局部跟踪器的目标框的交并 比,跟踪可分为理想、稳定和不稳定3种状态。局部 跟踪器和全局跟踪器的目标框的 IoU 为 0.5 时,跟 踪处于理想状态;大于某阈值时,跟踪处于稳定状态;小于该阈值时,跟踪处于不稳定状态。

不稳定状态下,综合考虑以下因素对跟踪器进 行分析:1)当前跟踪框和初始跟踪框的主色相关 度;2)响应图最高响应值;3)前一帧跟踪框中心到 当前跟踪框中心的移动距离。提取当前目标框的主 色,与初始帧的目标主色进行相关性计算,主色相关 度越低,说明跟踪器越有可能跟丢到非目标球员;跟 踪器的响应图最高响应值越小,在某种程度上说明 跟踪器的可靠性越低;跟踪框的移动距离大于一定 的阈值,说明跟踪器在当前帧很有可能突然发生跟 踪漂移。定位修改过程步骤为:

1) 计算修正前全局跟踪框 (x_g, y_g, w_g, h_g) 和局 部跟踪框 (x_p, y_p, w_p, h_p) 的 IoU。如果满足 1 > IoU $\geq \theta_1$, 进入步骤 5); 如果不满足, 进入步骤 2)。

2) 获取相关参数: Move_g, Con_g, CC_g, Move_p, Con_p, CC_p。Move 代表上一帧目标框中心到当前目 标框中心的移动距离, Con 代表跟踪器响应图的最 大响应值, CC 代表当前目标框和其初始帧目标框 的主色相关性,下标 g 代表全局跟踪器,下标 p 代表 局部跟踪器。

3)如果满足 $Move_g > \theta_2$, $Con_p > Con_g \times \theta_3$, $CC_p - CC_g > \theta_4 \equiv \uparrow$ 条件中的任何一个,则判定全 局跟踪器失效,局部跟踪器有效,全局跟踪框被修正 为 $(x_p, y_p, w_p, 2h_p)$,进入步骤 5);如果不满足,进 入步骤 4)。

4)如果满足 $Move_p > \theta_2$, $Con_g < Con_p \times \theta_3$, $CC_g - CC_p > \theta_4 \equiv \uparrow$ 条件中的任何一个,则判定局 部跟踪器失效,全局跟踪器有效,局部跟踪框被修正 为 $(x_g, y_g, w_g, h_g/2)$ 。

5) 输出修正后的全局跟踪框 (x_g, y_g, w_g, h_g) 和 局部跟踪框 (x_p, y_p, w_p, h_p) 。

定位修正前,局部跟踪框为 (x_p, y_p, w_p, h_p) ,全局跟踪框为 (x_g, y_g, w_g, h_g) ,如果算法认为全局跟踪器有效、局部跟踪器失效,局部跟踪器的跟踪结果则会被修正为 $(x_g, y_g, w_g, h_g/2)$;反之,全局跟踪器的跟踪结果则会被修正为 $(x_p, y_p, w_p, 2h_p)$ 。最终, 全局跟踪器的跟踪结果将作为输出的跟踪结果用于跟踪器性能分析。

2.3.4 模型更新

无论是全局跟踪器还是局部跟踪器,初始帧 的目标主色绝对可靠,更新操作可能会造成目标 主色被污染,故目标主色不更新;干扰项感知的颜 色模型中,目标球员的颜色偶尔受光照影响或者 部分被其他球员遮挡产生变化,且目标球员周围 的干扰项也会发生颜色变化,故对颜色模型采用 逐帧更新的策略以适应目标和干扰项在跟踪过程 中的颜色变化;目标感知的深度模型中,在得到目 标激活向量和模板分支上输出的目标感知的深度 特征后,目标激活向量和模板分支就不再逐帧计 算。全局跟踪器和局部跟踪器都具有一个固定不 变的深度模型和一个逐帧更新的颜色模型,这种 模型更新策略使得算法既能捕捉目标在跟踪过程 中发生的外观变化,又能防止因连续更新造成的 算法退化问题。

颜色模型的更新情况可以分为以下3种:

1)当前帧全局跟踪器和局部跟踪器均有效。 该情况下,对于全局跟踪器,假设当前帧为第t帧, η 为更新率, $P_{t-1}(x \in T | b_x)$ 为更新前的颜色模型, $P(x \in T | b_x)$ 为学习到的当前帧的颜色模型, $P_t(x \in T | b_x)$ 为更新后的颜色模型,使用线性插 值的方法对其颜色模型进行更新,具体为

$$P_{\iota}(x \in \mathbf{T} \mid b_{x}) =$$

$$\eta P(x \in \mathbf{T} \mid b_{x}) + (1 - \eta) P_{\iota-1}(x \in \mathbf{T} \mid b_{x})$$

(19)

对于局部跟踪器,采用和全局跟踪器相同的方 式对其颜色模型进行更新。

2)当前帧全局跟踪器失效、局部跟踪器有效。 该情况下,全局跟踪器被判定为失效跟踪器,说明前 若干帧的颜色模型很有可能被污染,为避免其颜色 模型产生错误累积,利用修正后的全局跟踪框重新

中国图象图形学报 JOURNAL OF IMAGE AND GRAPHICS

对于局部跟踪器,采用式(19)对其颜色模型进行更新。

3)当前帧全局跟踪器有效、局部跟踪器失效。 该情况下,对于全局跟踪器,采用式(19)对其颜色 模型进行更新;对于局部跟踪器,采用式(20)对其 颜色模型进行更新。

3 实 验

3.1 数据集

实验采用的数据集为 Soccer DatasetV2,该数据 集由 Yu 等人(2018)提出的 Soccer Dataset 改进而 来。Soccer Dataset 采用固定目标框对目标球员进行 标注,这种标注方式较简单,但是得到的目标框不够 精准。为获得更准确的实验结果,Soccer DatasetV2 对标注方式进行改进,摒弃固定目标框,采用精确的 目标框对目标球员进行标注。Soccer DatasetV2 具 体分布如表1所示。

Soccer DatasetV2 序列均来自远镜头中的连续 图片帧,共包含 80 个彩色图像序列,平均每个图像 序列 249 帧,最短序列 101 帧,最长序列 600 帧,共 19 908 个精准目标框。根据目标球员被遮挡的情况,Soccer DatasetV2 划分为4 类场景:第1 类场景 中目标球员没有被任何物体遮挡;第2 类场景中目 标球员被同队球员遮挡;第3 类场景中目标球员被 异队队员遮挡;第4 类场景中目标球员同时被多个 球员遮挡。

表 1 Soccer DatasetV2 数据集分布 Table 1 Distribution of Soccer DatasetV2

场景类型	图像序列编号	帧数
无遮挡场景	$1 \sim 20$	5 385
同队队员遮挡场景	$21 \sim 40$	3 954
异队队员遮挡场景	$41 \sim 60$	5 060
多遮挡场景	$61 \sim 80$	5 509

3.2 评价指标

评价跟踪算法一般从跟踪算法的跟踪准确率和 跟踪算法的跟踪速度两个方面进行考虑。 跟踪速度采用每秒跟踪帧数(frames per second, FPS)作为评价指标。

跟踪准确度采用平均有效重叠率(average valid overlap, AVO)作为跟踪准确度方面的评价指标。计算为

$$AVO = \frac{1}{M} \sum_{m=1}^{M} AVO_m$$
 (21)

式中, M 为足球序列的个数, AVO_m 为第 m 个序列的 平均有效重叠率, AVO_m 计算为

$$AVO_{m} = \frac{1}{N_{m}} \sum_{i=1}^{N_{m}} \varphi_{m,i}$$
 (22)

式中, N_m 为第 m 个序列的总帧数, $\varphi_{m,t}$ 为第 m 个序 列的第 t 帧的重叠率, 计算为

$$\varphi_{m,t} = \frac{A_{m,t}^{G} \cap A_{m,t}^{\mathrm{T}}}{A_{m,t}^{G} \cup A_{m,t}^{\mathrm{T}}}$$
(23)

式中, $A_{m,t}^{c}$ 为第 m 个序列的第 t 帧人工标注的目标 框, $A_{m,t}^{T}$ 为第 m 个序列的第 t 帧的目标框, 当 $\varphi_{m,t}$ 小 于 0.5 时, 将其重置为 0, 即

$$\varphi_{m,t} = \begin{cases} \varphi_{m,t} & \varphi_{m,t} \ge 0.5\\ 0 & \pm \ell t \end{cases}$$
(24)

3.3 跟踪准确度对比分析

实验选取 ATOM (accurate tracking by overlap maximization) (Danelljan 等, 2019)、CFNet (correlation filter network) (Valmadre 等, 2017)、CN (color name) (Danelljan 等, 2014b)、DAT (distractor-aware tracker) (Possegger 等, 2015)、DCFNet (discriminant correlation filters network) (Wang 等, 2017)、DSST (discriminative scale space tracker) (Danelljan 等, 2019)、SiamFC(fully-convolutional siamese networks) (Bertinetto 等, 2016b)、SiamRPN ++ (siamese visual tracking with very deep networks) (Li 等, 2019)、 SiamRPNRes22(deeper and wider siamese networks) (Zhang 和 Peng, 2019)和 Staple(sum of template and pixel-wise learners)(Bertinetto 等, 2016a)作为对比 跟踪器。PAT-nopart 表示本文算法没有采用局部跟 踪器以及定位修正策略的结果。

各跟踪器在 Soccer DatasetV2 上的跟踪准确度 表现如表 2 所示。

3.3.1 局部跟踪器消融实验的表现

从实验结果中可以看出,在不采用局部跟踪器 以及定位修正模块时,本文算法的准确度表现不佳。 主要原因是在出现球员遮挡的情况下容易发生跟踪

Table 2	The AVO measurement on Soccer Dataset V2
Tabla 2	The AVO measurement on Soccar DatasetV2
表 2	各跟踪器在 Soccer DatasetV2 上的 AVO

跟踪器	AV01	AVO2	AVO3	AV04	平均 AVO
ATOM	0. 758 5	0.358 5	0.425 0	0.202 5	0.436 1
CFNet	0.632 0	0.3645	0.437 0	0.3505	0.446 0
CN	0.5195	0. 190 5	0.208 0	0.172 5	0.272 6
DAT	0.6945	0.356 0	0. 499 0	0.334 0	0. 470 9
DCFNet	0.688 5	0.455 5	0.4905	0.3755	0.5025
DSST	0.5835	0.265 0	0.2395	0.224 0	0.328 0
SiamFC	0.5970	0.264 5	0.3215	0.234 0	0.3543
SiamRPN ++	0.798 0	0. 471 0	0.4900	0.295 5	0. 513 6
SiamRPNRes22	0.5490	0.273 0	0.385 5	0. 196 0	0.3509
Staple	0.6215	0.352 0	0.451 0	0.363 0	0.4469
PAT-nopart	0.609	0. 197 5	0.217 5	0.085 5	0.2774
PAT	0.722 5	0.508 0	0.5645	0.446 5	0.5603
平均值	0.6479	0.338 0	0.394 1	0.273 3	0.413 3

注:加粗字体和斜体分别表示最优和次优结果。

漂移,而跟踪漂移多发生在整个序列前几帧,从而使整个视频序列跟踪失败,导致出现无效结果。PATnopart 与 PAT 的实验结果说明了球员跟踪的重大挑战在于相似球员之间的遮挡与干扰,在足球场景的球员跟踪算法中增加局部跟踪以及重新定位模块对整个算法有着巨大的提升作用。

3.3.2 无遮挡场景的表现

无遮挡场景是足球比赛视频中最简单的场景, 目标球员周围大多是单一的场地背景,该场景下目 标球员可能快速跑动,跟踪算法需要对目标快速运 动及时作出反应。实验结果显示,各跟踪器在无遮挡 场景下的跟踪准确度的排名依次为:SiamRPN ++ > ATOM > PAT > DAT > DCFNet > CFNet > Staple > SiamFC > DSST > SiamRPNRes22 > CN。各跟踪器在 无遮挡场景下的平均 AVO 得分为 0.651 3,其中, PAT 以 0.722 5 的单项得分排名第 3,且与排名第 1 的 SiamRPN ++ 相差 7.6%。

3.3.3 同队球员遮挡场景的表现

同队球员遮挡场景下,目标球员和同队球员非 常相似,需要跟踪算法具备极强的辨别能力。实验 结果显示,各跟踪器在同队遮挡场景下的跟踪准确 度的排名依次为:PAT > SiamRPN ++ > DCFNet > CFNet > ATOM > DAT > Staple > SiamRPNRes22 > DSST > SiamFC > CN。各跟踪器在同队球员遮挡场 景下的平均 AVO 得分为 0.350 8,其中,PAT 以 0.508 0 的单项得分排名第 1,且比排名第 2 的 SiamRPN ++ 高出 3.7%。

3.3.4 异队球员遮挡场景的表现

异队球员遮挡场景下,不同球队队员间颜色差 距较大,目标球员较容易辨认,但是跟踪算法需要有 良好的模型更新机制应对长时间的异队球员遮挡带 来的样本污染问题。实验结果显示,各跟踪器在异 队遮挡场景下的跟踪准确度的排名依次为:PAT > DAT > DCFNet > SiamRPN ++ > Staple > CFNet > ATOM > SiamRPNRes22 > SiamFC > DSST > CN。各 跟踪器在异队球员遮挡场景下的平均 AVO 得分为 0.410 1,其中,PAT 以 0.564 5 的单项得分排名第 1,且比排名第2 的 DAT 高出 6.6%。

3.3.5 综合场景的表现

实验结果显示,各跟踪器在所有场景下的跟踪准确度的排名依次为:PAT > SiamRPN ++ > DCFNet > DAT > Staple > CFNet > ATOM > SiamFC > SiamR-PNRes22 > DSST > CN。各跟踪器在所有场景下的平均 AVO 得分为 0.425 6,其中,PAT 以 0.560 3 的综合得分排名第 1,且比排名第 2 的 SiamRPN ++ 高出 4.7%。

3.3.6 结果分析

从各个跟踪器在各个场景下的 AVO 数值分布 看,4 个场景的跟踪难度从高到低依次为:多球员遮 挡场景 > 同队球员遮挡场景 > 异队球员遮挡场景 > 无遮挡场景。各个跟踪器在无遮挡场景下的表现均 不错。除了无遮挡场景,PAT 在其他 3 个场景下均 排名第 1,且相较于这 3 个场景下排名第 2 的算法 均有不同程度的提升。按照 PAT 在各个场景下相 对第 2 名的提升排名依次为:多球员遮挡场景 > 同 队球员遮挡场景 > 异队球员遮挡场景。这说明 PAT 在足球比赛视频中的简单场景下表现良好,且相比 其他各种类型的跟踪算法更能从容应对足球比赛视 频中的高难度跟踪场景。这得益于以下几点:

1)干扰项感知的颜色模型法能够轻松辨别目 标球员与异队球员,同时能够帮助辨别目标球员与 同队球员;

2)目标感知的深度特征有利于区分类内差异, 增强了算法对目标球员与其他球员的辨别能力; 3) 深度模型固定不更新, 使得 PAT 在目标球员 被遮挡后仍能够"记得"目标球员;

4) 跟踪算法出现漂移现象时, PAT 中的局部跟踪器及时与全局跟踪器协同工作, 修正目标定位。

3.4 跟踪速度对比分析

各跟踪算法在 Soccer DatasetV2 上的跟踪速度 如表 3 所示。实验结果表明,各跟踪器在跟踪速度 方面的排名依次为: CN > DSST > DAT > Staple > DCFNet > SiamRPNRes22 > SiamFC > ATOM > CFNet > PAT > SiamRPN ++。

表 3	各跟	踪器在	Socc	er D	atase	tV2	上的	FPS
Tabl	e 3	The FI	PS m	easui	emer	nt on	Soco	cer
DatasetV2 of each tracking algorithm								

算法	帧率/(帧/s)			
АТОМ	17.02			
CFNet	17.00			
CN	120. 92			
DAT	66. 97			
DCFNet	32. 94			
DSST	69. 73			
SiamFC	23.68			
SiamRPN + +	8.44			
SiamRPNRes22	23.88			
Staple	60. 92			
PAT	9.79			

注:加粗字体和斜体分别表示最优和次优结果。

干扰项感知的颜色模型除了要计算目标—背景 颜色模型,还要计算目标—干扰项颜色模型,故干扰 项感知的颜色模型的计算量是普通颜色模型的两 倍。目标感知的深度模型需要在初始帧在线训练— 个回归网络。局部跟踪器的加入使得 PAT 的计算 量增加近一倍。上述一系列操作增加了算法的计算 复杂度,导致该算法无法达到实时性要求。

4 结 论

针对足球比赛视频这一特定跟踪场景,通过融 合干扰项感知的颜色模型和目标感知的深度模型, 提出了一种球员感知的跟踪算法。颜色模型方面, 提取目标球员、背景以及干扰项的颜色直方图,利用

贝叶斯公式计算搜索图像中各像素点属于目标的似 然概率。深度模型方面,利用回归网络从深度特征 中选择与当前跟踪目标相关的特征通道,利用孪生 网络计算搜索图像与目标图像的相似度。此外,针 对跟踪漂移问题,提出使用全局跟踪器和局部跟踪 器分别跟踪目标整体和目标上半身,当两个跟踪结 果的距离大于一定的阈值时,通过主色相关度、最高 响应值和移动距离等参数判断两个跟踪器的有效 性,利用有效跟踪器的跟踪结果修正失效跟踪器的 跟踪结果。实验结果表明,在相似球员干扰和遮挡 等挑战下,球员感知跟踪算法的跟踪效果明显优于 其他算法。但是由于干扰项感知的颜色模型计算、 目标感知的深度模型计算和局部跟踪器的加入等增 加了算法计算复杂度,跟踪速度较慢,未来工作将聚 焦在算法优化,减少重复计算,提高算法的跟踪 效率。

参考文献(References)

- Bastanfard A, Jafari S and Amirkhani D. 2019. Improving tracking soccer players in shaded playfield video//Proceedings of the 5th Iranian Conference on Signal Processing and Intelligent Systems. Shahrood, Iran; IEEE; 1-8 [DOI: 10.1109/ICSPIS48872.2019.9066103]
- Baysal S and Duygulu P. 2016. Sentioscope: a soccer player tracking system using model field particles. IEEE Transactions on Circuits and Systems for Video Technology, 26(7): 1350-1362 [DOI: 10. 1109/TCSVT. 2015. 2455713]
- Bertinetto L, Valmadre J, Golodetz S, Miksik O and Torr P H S. 2016a. Staple: complementary learners for real-time tracking//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 1401-1409 [DOI: 10.1109/CVPR. 2016.156]
- Bertinetto L, Valmadre J, Henriques J F, Vedaldi A and Torr P H S. 2016b. Fully-convolutional Siamese networks for object tracking// Proceedings of the European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 850-865 [DOI: 10.1007/ 978-3-319-48881-3_56]
- Danelljan M, Bhat G, Khan F S and Felsberg M. 2019. ATOM: accurate tracking by overlap maximization//Proceedings of 2019 IEEE/ CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4655-4664 [DOI: 10.1109/CVPR. 2019. 00479]
- Danelljan M, Häger G, Khan F S and Felsberg M. 2014a. Accurate scale estimation for robust visual tracking//Proceedings of British Machine Vision Conference. Guildford, Surrey UK: British Machine Vision Association Press: #79 [DOI: 10.5244/C.28.65]

冯思佳,宋子恺,于俊清,何云峰,管涛 / 足球视频球员感知跟踪算法

- Danelljan M, Khan F S, Felsberg M and Van De Weijer J. 2014b. Adaptive color attributes for real-time visual tracking//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA; IEEE: 1090-1097 [DOI: 10.1109/CVPR. 2014.143]
- Fu L. 2015. A Method for Multi-Target Tracking in Soccer Videos. Shijiazhuang: Hebei University of Technology (付龙. 2015. 足球视频 中多目标跟踪算法研究. 石家庄:河北工业大学)
- Henriques J F, Caseiro R, Martins P and Batista J. 2015. High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3): 583-596 [doi: 10.1109/TPAMI.2014.2345390]
- Heydari M and Moghadam A M E. 2012. An MLP-based player detection and tracking in broadcast soccer video//Proceedings of 2012 International Conference of Robotics and Artificial Intelligence. Rawalpindi, Pakistan; IEEE; 195-199 [DOI; 10. 1109/ICRAI. 2012. 6413398]
- Kataoka H and Aoki Y. 2011. Football players and ball trajectories projection from single camera's image//Proceedings of the 17th Korea-Japan Joint Workshop on Frontiers of Computer Vision. Ulsan, Korea (South): IEEE: 1-4 [DOI: 10.1109/FCV.2011.5739712]
- Li B, Wu W, Wang Q, Zhang F Y, Xing J L and Yan J J. 2019. Siam-RPN ++ : evolution of Siamese visual tracking with very deep networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4277-4286 [DOI: 10.1109/CVPR.2019.00441]
- Li H P and Flierl M. 2012. Sift-based multi-view cooperative tracking for soccer video//Proceedings of 2012 IEEE International Conference on Acoustics, Speech and Signal Processing. Kyoto, Japan: IEEE: 1001-1004 [DOI: 10.1109/ICASSP.2012.6288054]
- Liu J C, Carr P, Collins R T and Liu Y X. 2013. Tracking sports players with context-conditioned motion models//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE: 1830-1837 [DOI: 10.1109/CVPR.2013.239]
- Lou N, He N Z and Shi B C. 2007. Detection and tracking in soccer video sequences. Computer Engineering and Applications, 43(2): 227-230 (娄娜,何南忠,施保昌. 2007. 足球视频中的目标检测与跟踪. 计算机工程与应用,43(2): 227-230)[DOI: 10. 3321/j.issn:1002-8331.2007.02.067]
- Lu W L, Ting J A, Murphy K P and Little J J. 2011. Identifying players in broadcast sports videos using conditional random fields//Proceedings of the Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE: 3249-3256 [DOI: 10.1109/ CVPR.2011.5995562]
- Ma Y J, Feng S and Wang Y B. 2018. Research on player tracking algorithm based on deep learning. Journal of Communication University of China (Science and Technology), 25(3): 60-64 (马月洁, 冯爽, 王永滨. 2018. 基于深度学习的足球球员跟踪算法研究. 中国传媒大学学报(自然科学版), 25(3): 60-64) [DOI: 10.

16196/j. cnki. issn. 1673-4793. 2018. 03. 009]

- Mazzeo P L, Spagnolo P, Leo M and D'Orazio T. 2008. Visual players detection and tracking in soccer matches//Proceedings of the 5th International Conference on Advanced Video and Signal Based Surveillance. Santa Fe, USA: IEEE: 326-333 [DOI: 10.1109/ AVSS.2008.33]
- Morais E, Goldenstein S, Ferreira A and Rocha A. 2012. Automatic tracking of indoor soccer players using videos from multiple cameras//Proceedings of the 25th SIBGRAPI Conference on Graphics, Patterns and Images. Ouro Preto, Brazil: IEEE: 174-181 [DOI: 10.1109/SIBGRAPI.2012.32]
- Najafzadeh N, Fotouhi M and Kasaei S. 2015. Multiple soccer players tracking//Proceedings of the International Symposium on Artificial Intelligence and Signal Processing. Mashhad, Iran: IEEE: 310-315 [DOI: 10.1109/AISP.2015.7123503]
- Possegger H, Mauthner T and Bischof H. 2015. In defense of colorbased model-free tracking//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 2113-2120 [DOI: 10.1109/CVPR.2015.7298823]
- Seo Y, Choi S, Kim H and Hong K S. 1997. Where are the ball and players? Soccer game analysis with color-based tracking and image mosaick//Proceedings of the Image Analysis and Processing. Florence, Italy: Springer: 196-203 [DOI: 10.1007/3-540-63508-4_ 123]
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2020-08-21]. https://arxiv.org/pdf/1409.1556v4.pdf
- Valmadre J, Bertinetto L, Henriques J, Vedaldi A and Torr P H S. 2017. End-to-end representation learning for correlation filter based tracking//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5000-5008 [DOI: 10.1109/CVPR.2017.531]
- Wang N Y and Yeung D Y. 2013. Learning a deep compact image representation for visual tracking//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, USA: Curran Associates Inc.: 809-817
- Wang Q, Gao J, Xing J L, Zhang M D and Hu W M. 2017. DCFNet: discriminant correlation filters network for visual tracking [EB/ OL]. [2020-08-21]. https://arxiv.org/pdf/1704.04057.pdf
- Wang X. 2017. Research on Player Tracking Algorithm in Soccer Video. Wuhan: Huazhong University of Science and Technology (王勋. 2017. 足球视频中球员跟踪算法研究. 武汉:华中科技大学)
- Yu J Q, Lei A P, Song Z K, Wang T T, Cai H Y and Feng N. 2018. Comprehensive dataset of broadcast soccer videos//Proceedings of 2018 IEEE Conference on Multimedia Information Processing and Retrieval. Miami, USA: IEEE: 418-423 [DOI: 10.1109/MIPR. 2018.00090]
- Zhang Z P and Peng H W. 2019. Deeper and wider Siamese networks for real-time visual tracking//Proceedings of 2019 IEEE/CVF Confer-

ence on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4586-4595 [DOI: 10.1109/CVPR.2019.00472]

作者简介

中国图象图形学报

JOURNAL OF IMAGE AND GRAPHICS



冯思佳,1994年生,女,硕士研究生,主要研 究方向为单目标视觉跟踪。 E-mail: 1140633868@qq.com



于俊清,通信作者,男,教授,主要研究方向 为基于内容的视频分析、索引与检索、多核 计算与流编译、视频情感计算、网络安全与 大数据处理。 E-mail: yjqing@hust.edu.cn

宋子恺,男,博士研究生,主要研究方向为单目标视觉跟踪、 图像特征编码。E-mail: skyesong@ hust. edu. cn 何云峰,男,副教授,主要研究方向为数字媒体处理与检索、 计算机动画与影视特技。E-mail: yfhe@ hust. edu. cn 管涛,男,教授,主要研究方向为增强现实、虚拟现实、无人机 视觉导航、避障技术、全自动3维建模技术、面向嵌入式设备 的计算机视觉技术。E-mail: qd_gt@ 126. com