

# 体育视频中的运动员检测与分割\*

姚沁汝<sup>1</sup>, 唐九飞<sup>2</sup>, 于俊清<sup>1,2</sup>, 王赠凯<sup>1</sup>

<sup>1</sup>(华中科技大学 计算机科学与技术学院, 湖北 武汉 430074)

<sup>2</sup>(华中科技大学 网络与计算中心, 湖北 武汉 430074)

通讯作者: 唐九飞, E-mail: tangjiufei@hust.edu.cn

**摘要:** 运动员的动作行为分析是体育视频高级语义分析的直接途径, 检测和分割视频中的运动员是分析运动员的动作行为的基础. 利用体育视频的领域规则和中层特征块的性质, 提出半监督的方法挖掘运动员的中层特征块, 针对不同类型的镜头分别训练基于中层特征块的运动员检测分类器, 实现运动员检测. 利用运动员检测结果标记超像素, 结合 Grab Cut 分割算法实现运动员分割. 实验结果表明, 基于中层特征块的运动员区域检测算法能够快速挖掘训练所需样本, 从而训练得到检测分类器, 检测结果具有较高的准确度, 获得的运动员区域能够有效用于运动员分割, 简化了分割计算过程.

**关键词:** 视频分析; 目标检测; 目标分割; 中层特征块; 超像素

中文引用格式: 姚沁汝, 唐九飞, 于俊清, 王赠凯. 体育视频中的运动员检测与分割. 软件学报, 2015, 26(Suppl. (2)): 155-164. <http://www.jos.org.cn/1000-9825/15026.htm>

英文引用格式: Yao QR, Tang JF, Yu JQ, Wang ZK. Player detection and segmentation in sports video. Ruan Jian Xue Bao/ Journal of Software, 2015, 26(Suppl. (2)): 155-164 (in Chinese). <http://www.jos.org.cn/1000-9825/15026.htm>

## Player Detection and Segmentation in Sports Video

YAO Qin-Ru<sup>1</sup>, TANG Jiu-Fei<sup>2</sup>, YU Jun-Qing<sup>1,2</sup>, WANG Zeng-Kai<sup>1</sup>

<sup>1</sup>(School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

<sup>2</sup>(Center of Network and Computation, Huazhong University of Science and Technology, Wuhan 430074, China)

**Abstract:** Action and behavior analysis of players is a direct method of high-level semantic analysis or highlight annotation in sports video. Accurate detection and segmentation of players is the key technology of this method. Employing domain knowledge and characteristics of mid-level feature patch in sports video, a semi-supervised algorithm is proposed to discover the mid-level feature patch and train the player detector for different types of video shots. The detection result is used to label the superpixel, and then player segmentation is accomplished by Grab Cut segmentation algorithm. Experimental results show that the mid-level feature patch based player detector is convenient to train and achieves high detection accuracy. The detected player regions can be used to segment the players effectively, and hence the computation procedure of player segmentation is simplified.

**Key words:** video analysis; object detection; object segmentation; mid-level feature patch; superpixel

运动员的行为与动作分析对体育视频高级语义分析如精彩事件检测的具有重要作用. 将运动员从背景中分割出来, 能够消除背景的干扰, 降低运动员行为动作分析的难度. 运动员在体育视频中一般处于运动状态, 检测与分割体育视频中的运动员属于视频运动对象的检测分割问题. 从视频连续帧间具有的空间连续性及运动对象自身特点的不同角度考虑, 视频中运动对象的检测分割方法可以分为基于背景建模分割前景背景、基于区分前景与背景的相对运动检测分割前景和基于前景的特点检测分割前景这 3 类方法.

基于背景建模方法一般用于摄像头静止的情况, 通过构建视频中的背景预测和标记待检测图像的像素点

\* 基金项目: 国家自然科学基金(61173114); 武汉市应用基础研究计划(2014010101010027)

收稿时间: 2014-06-20; 定稿时间: 2014-08-20

为前景或背景像素点.预测标记像素点的方法与构建背景信息的方法息息相关,主要分为两类:一类存储背景图像,将背景图像与待测图像作差,对作差结果使用阈值来预测标记像素点<sup>[1]</sup>;另一类通过统计若干样本图像建立背景像素点的模型,判断像素点是否属于对应背景像素点的模型,从而标记像素点的属性<sup>[2-4]</sup>.基于区分前景与背景运动的方法不限制摄像机是否运动,这种方法的出发点是图形中背景的运动一般由摄像机的运动引起的,而图像中前景的运动是由前景本身的运动与摄像机的运动共同作用产生,所以前景与背景之间存在相对运动,通过检测相对运动从而判断标记出前景运动的像素点.通过跟踪像素点得到像素点的运动轨迹矩阵,由于背景像素点的运动由摄像机移动产生,所以背景像素点的轨迹矩阵是低秩的,而前景像素点由摄像机和对象各自本身的运动产生,所以运动轨迹矩阵不会是低秩.通过这一性质判断背景与前景像素点,再根据像素点的空间连续性进一步分割前景对象<sup>[5-7]</sup>.基于前景特点检测分割前景对象的方法就是目标检测与目标分割的方法.目标检测与分割是计算机视觉领域研究的基础问题<sup>[8,9]</sup>.选择能够描述表征目标的特征表示目标是目标检测与分割的第1步.不同特征表示对目标的描述能力和抗干扰能力不同<sup>[10-13]</sup>.在实现目标检测的基础上,结合图像的空间信息和相关的先验知识实现目标分割<sup>[14,15]</sup>.

本文的主要贡献有以下3点:(1) 基于中层特征块的性质<sup>[16]</sup>,提出了一种半监督挖掘中层特征块的方法,用于训练运动员检测分类器;(2) 提出了一种利用基于中层特征块训练的检测分类器检测运动员的方法;(3) 结合运动员检测结果,以超像素为基本单位标记像素,应用 Grab Cut 算法<sup>[17]</sup>实现运动员分割.

## 1 体育视频特点分析与算法概述

### 1.1 体育视频特点

体育视频不同类型的镜头中图像有着各自的特点.在特写镜头中,运动员只有半身像,因为摄像机的焦点在运动员身上,相对于背景中的观众,运动员的轮廓比观众轮廓要大;中镜头和远镜头中运动员都是全身像,不同的是中镜头中的运动员比远镜头中的运动员轮廓要大,且中镜头中的背景较为复杂;远镜头中背景主要是比赛场地、少量广告牌和观众.基于以上特点,按照以下3个原则训练不同镜头类型的运动员分类器:

- (1) 在远镜头中,先消除场地色再检测运动员,以降低分类器的复杂度,加快检测速度;
- (2) 根据不同镜头中运动员大小不同的特点,3种镜头下的训练分类器的中层特征块尺寸不同;
- (3) 各类型镜头中图像的运动员与背景最大的区别在于块的轮廓和结构,使用 HOG 特征作为图像块的特征.

### 1.2 运动员的检测与分割算法概述

在体育视频拍摄过程中,摄像机随着运动员的运动而移动,导播会运用不同的镜头来生动地展现比赛场景.因此,基于背景建模的方法不适合用来检测分割体育视频中的运动员.基于区分背景前景运动检测分割运动员的方法不受上述因素的影响,但这类方法用于体育视频分析存在以下问题:第一,像素点的运动信息的计算开销很大;第二,像素点的运动轨迹生成需要跟踪多帧图像像素点,然后才能建立需要的模型检测分割运动员,即不能根据高级语义分析的需要,实时检测分割指定图像中的运动员;第三,在同一个视频中,不同镜头中的内容不一定连续,前一个镜头建立的各种背景与前景模型不一定能够适应其后镜头的场景,这造成了需要根据镜头变换而不断重新建立模型来检测新镜头的内容,加大了计算量和计算复杂度.所以,基于时空联合的方法不适合用于体育视频中运动员的检测和分割.基于目标特征的检测分割方法通过选择合适的特征表示分割对象,收集样本训练目标分类器,实现视频中运动员的检测和分割,适用于体育视频.

体育视频中不同镜头的特点为检测和分割运动员提供线索.体育视频运动员的检测与分割算法框架如图1所示,通过镜头分割<sup>[18]</sup>将视频分割成镜头序列,识别出镜头类型<sup>[19]</sup>,再根据不同镜头的特点分别实现运动员的检测与分割:在远镜头中的场地外不会存在运动员,利用这一特点在检测与分割之前进行场地检测以缩小检测区域;对于中镜头,利用场地颜色信息进行运动员分割;对于特写镜头,直接进行运动员的检测与分割.

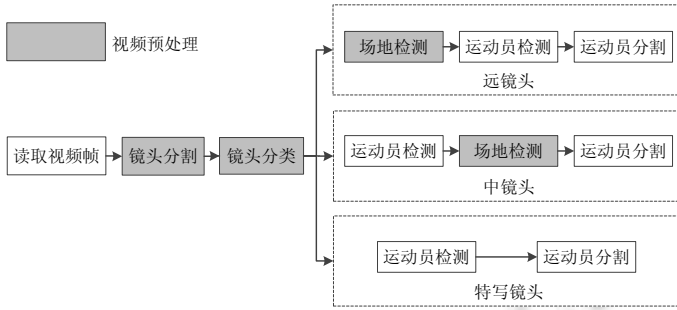


Fig.1 Framework of player detection and segmentation

图1 体育视频中运动员的检测与分割算法框图

## 2 体育视频中运动员的检测

### 2.1 基于中层特征块的运动员检测分类器训练过程

Saurabh 等人<sup>[16]</sup>最先提出物体的中层特征块的概念,物体的中层特征块能够代表目标对象并将其与其他物体区分出来,即中层特征块是部件级或部件级以上的物体特征概念.文献[17]总结了中层特征块具有的两个特征:(1) 代表性,即在目标物体出现的同时,该部件出现的频率要足够高;(2) 区分性,即能够区分目标物体与非目标物件.在体育视频中,运动员与背景总是同时存在,这种情况下,实现无监督挖掘运动员中层特征块过于困难.本文提出一种半监督挖掘运动员中层特征块的方法,并训练运动员的检测分类器.

在特写镜头中,运动员只有半身像,其形态种类不多,分割出的含有运动员不同部分的块的种类也不多.在中镜头和远镜头中,运动员在图像中所占比例较小,尤其在足球、网球和羽毛球的远镜头中,大部分运动员都能使用一个不大的块完整分割.中镜头和远镜头中的图像分割出的含有运动员的差异大的块数也不多.基于这一特点,在同一个视频中人工初始选定特征块,根据欧氏距离计算与初始特征块最相似的若干块以及最不相似的若干块作为初始样本集,训练初始粗糙的检测分类器.接下来进行人工监督迭代优化检测分类器的过程:迭代地使用分类器对所有块分类;用 K-means 对正、负结果的图片进行聚类,查看每个聚类结果对应的分类是否有误,如果有误,则把错误的类加入到正确的样本,如此扩展正、负样本集;下一次迭代的开始重新提取样本集特征并训练新的分类器,优化了分类器;迭代检测、聚类和更新优化分类器,直到分类器对一定数量新视频的图片集也正确分类为止.具体过程如图 2 所示.其中,为了使图像间相似度的计算更为准确,每次加入的图片集为同一个视频的,当检测收敛时再加入新的视频图像集.

### 2.2 运动员检测过程

根据不同类型体育视频和不同镜头的运动员特点,分割不同尺寸的特征块.使用与特征块尺寸相同的窗口和一定的步长遍历图像,检测图像中的运动员块.为了防止分类器分类错误引起的运动员误检情况,在图像遍历

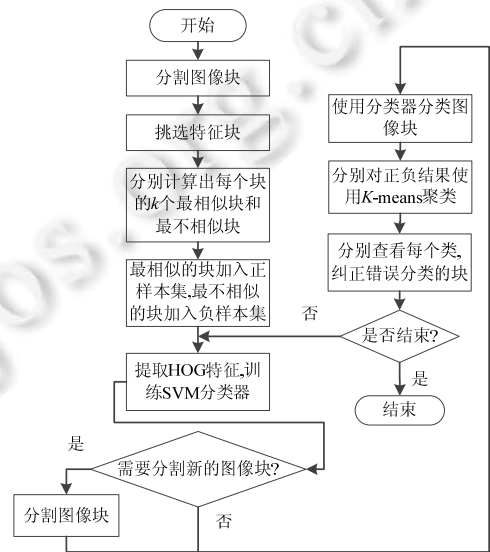


Fig.2 Training procedure of player classifier

图2 运动员分类器训练流程图

结束后需要有确认检测结果的机制。

运动员中层特征块是部件级特征,检测到的运动员区域由多个块组成,且块与块之间会有重叠区域.而分类器分类出错是小概率事件,因此误检的块与块之间存在重叠的可能性很小.利用这一特点,给检测窗口赋予权重(如图 3 所示),并叠加检测窗口的权重,权重叠加的结果是块重叠越多的区域的值越高,通过阈值化处理丢弃离散的误检块.这样,过程简单且能够减少误检情况的出现.

运动员的检测流程如图 4 所示,采用与特征块大小相同的窗口和设定步长来遍历图像,初始化一个大小与图像相同,所有像素灰度值为 0 的灰度图像 mask 用来存储检测图像结果.对于每次窗口检测结果为运动员块的时候,将该块的权重与 mask 对应的块相加存储到 mask 中.为了防止运动员分类器分类错误造成的运动员误检,遍历检测图像结束后对存储检测结果的 mask 进行阈值化处理,获取运动员的位置和区域信息.

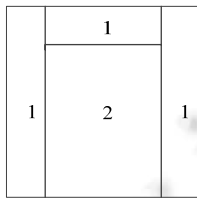


Fig.3 Weights of player patches  
图 3 运动员的图像块权重

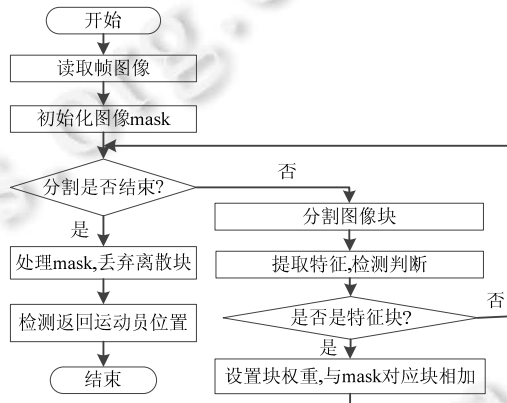


Fig.4 Flowchart of player detection  
图 4 运动员检测流程图

### 3 体育视频中运动员的分割

基于中层特征块的运动员检测结果能够提供两种运动员的表示:矩形框和多边形区域表示.图 5 是特写镜头中基于中层特征块的运动员检测结果.其中,图 5(a)是检测到的运动员中层特征块;图 5(b)是运动员检测结果的矩形框表示.在多边形区域表示的运动员区域中,多边形的中间部分属于运动员像素,边缘区域属于背景像素,利用这些线索标记多边形区域中的像素是属于运动员像素还是背景像素,并进一步用于运动员分割.

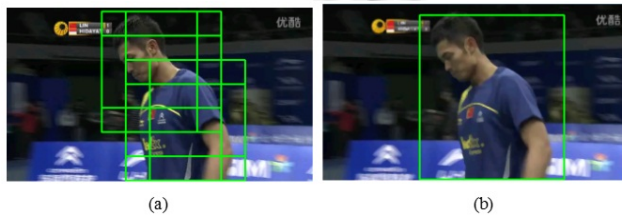


Fig.5 Player detection result of close-up shot

图 5 特写镜头中运动员检测结果

#### 3.1 超像素与Grab Cut分割算法

##### 1. 超像素

超像素的概念首先由 Ren 等人在文献[20]中提出.超像素指空间位置上相邻并且纹理、颜色和亮度等特征相似的像素点集合.与以像素为基本分割单元的分割算法相比,以超像素为基本单元分割图像,既能保持图像中轮廓信息,又避免大量像素的冗余计算.目前已经有超像素分割算法,主要分为基于图论和基于梯度下降两

类.基于图论分割超像素算法的首要步骤是根据图像的像素信息构建一个图来初始化分割.构建图的方法是:图的顶点是图像中每个像素点,图的边大小是对应两个像素点的相似度<sup>[21,22]</sup>.分割过程就是最小化分割代价函数的过程.基于梯度下降的超像素分割算法一般都以对像素作粗糙聚类初始分割,然后基于梯度下降的思路迭代优化聚类直至收敛以达到最优的的分割效果<sup>[23,24]</sup>.

SLIC(simple linear iterative clustering)算法<sup>[25]</sup>是 Aravind 等人提出的一种简单线性迭代聚类分割超像素方法,使用像素的三维 CIELAB 颜色空间值和二维空间坐标构建像素的 5 维特征,对图像中的像素点迭代局部聚类直至收敛.该算法速度比较快,可以人为指定超像素大小,生成紧凑和近似均匀的超像素.采用 SLIC 超像素分割方法的原因主要有两点:第一,SLIC 超像素分割方法可以通过指定超像素个数来控制超像素分割粒度.运动员检测分割算法是分镜头类型实现,不同镜头类型中的运动员大小不一样,运动员区域内的超像素粒度也不一样,需要区分不同镜头中的超像素粒度;第二,SLIC 超像素分割算法简单,速度较快.

## 2. Grab Cut 分割算法

Grab Cut 算法是 Rother 等人在文献[17]中提出的一种图像分割算法.该算法利用图像中的颜色、纹理和少量用户标记信息分割图像中的物体.采用多通道混合高斯模型描述图像的前景背景的颜色信息,根据交互信息中的像素初始标记估计各类 GMM(Gauss mixture model,简称 GMM)参数,通过迭代最小化分割能量标记像素类别以达到最优分割.用户的交互信息是以指定前景框的形式输入,表示用户对图像中所有像素的初始标记:前景框外的像素标记为背景像素,框内的像素标记为可能前景的像素.用户还可以输入更详细的标记信息,将前景框内的像素点指定为前景、背景、可能前景或可能背景像素 4 种中的 1 种.在此基础上初始化分割:对每个类别的像素分别使用 K-means 算法聚类,设定类的个数为 K,即对应各类 GMM 中的 K 个高斯分量,每个类中的像素是属于对应高斯分量中的像素样本.GMM 的均值和协方差通过样本像素的像素值估计得到,各个高斯分量的权值是属于对应高斯分量的像素个数与对应类别的总像素个数的比值;迭代地根据 GMM 参数更新像素标记,利用 max flow/min cut 算法进行图像分割,重新标记每个像素并重估 GMM 参数,直至收敛.

## 3.2 运动员分割过程

结合运动员检测的运动员区域信息结果,以超像素为标记单位,根据超像素的大小位于运动员区域的位置标记超像素的属性,标记结果作为 Grab Cut 分割算法的初始化信息.运动员分割算法的具体流程如图 6 所示:分割运动员之前对图像分割超像素,然后利用运动员的多边形区域表示计算超像素的分类.遍历图像统计每个超像素的总像素个数和每个超像素在运动员多边形区域内的像素个数,计算每个超像素的在多边形区域内的像素数与像素总数的比值.使用两个阈值  $T1, T2$  用来判断标记超像素,将超像素标记为 3 类:前景超像素、可能前景超像素和可能背景超像素.当超像素的比值  $\geq T1$  时,则该超像素是前景超像素;当  $T1 \leq$  像素比值  $\leq T2$  时,该超像素是可能前景超像素;当超像素比值  $\leq T2$  时,该超像素属于可能背景超像素.阈值  $T1$  和  $T2$  的设置需要根据中层特征块的选择设置,如果中层特征块包含的背景比较多,则  $T1$  与  $T2$  的值设置相对较大一些;反之则阈值  $T1$  与  $T2$  的值设置较小一些.将运动员所在位置的矩形框和前景背景标记信息作为交互信息传递给 Grab Cut 函数,最终得到单个运动员的最优分割.Grab Cut 分割运动员的结果存储在使用 0 和 1 标记的单通道 mask,在一帧图像中含有多

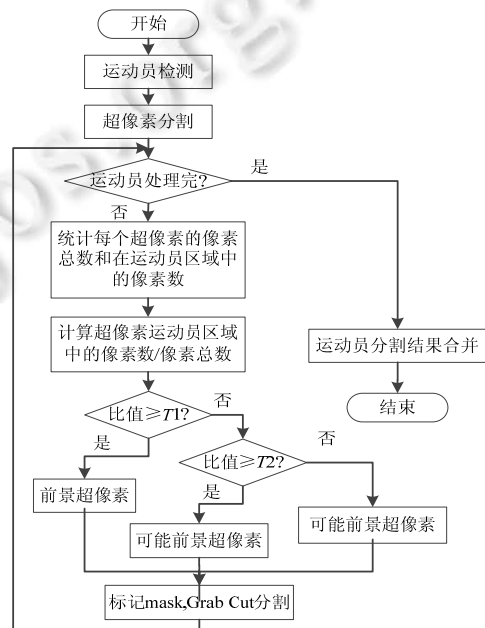


Fig.6 Flowchart of player segmentation

图 6 运动员分割流程图

个运动员的情况下,需要合并每次分割运动员的结果 mask,即将所有结果 mask 作或操作,得到最终运动员检测分割结果.

### 4 实验结果与分析

#### 4.1 体育视频中运动员的检测

##### 1. 特写镜头运动员检测

选取视频帧尺寸为 512×288 的足球和羽毛球视频作为采样和测试的视频.由于特写镜头中的图像细节被放大,可能存在半身的观众,但观众的大小与运动员的大小存在差别.所以,特写镜头中检测窗口(即中层特征块尺寸)一般设置较大尺寸,此时用来描述运动员的中层特征块是部件级特征.实验设置中层特征块的尺寸为 160×160,步长为 40×40.检测结果如图 7 所示.其中,前两行是羽毛球视频特写镜头的运动员检测结果,后两行是足球视频特写镜头的运动员检测结果;图 7(a)中的矩形框表示检测到的运动员中层特征块,图 7(b)中的矩形框表示最终的运动员检测结果.

##### 2. 中镜头运动员检测

选取视频帧尺寸为 512×288 的足球和羽毛球视频作为采样和测试的视频.中镜头中的运动员是全身像,运动员的大小随着运动员的动作发生变化,故中层特征块的大小应能适应运动员的形变.中镜头的运动员中层特征块仍是部件级特征,但不区分运动员的具体部位.实验设置中镜头运动员中层特征块尺寸为 80×160,步长为 20×40.检测结果如图 8 所示.其中,前两行是足球视频中镜头的运动员检测结果,后两行是羽毛球视频中镜头的运动员检测结果;图 8(a)中的矩形框表示检测到的运动员中层特征块,图 8(b)中的矩形框表示最终的运动员检测结果.

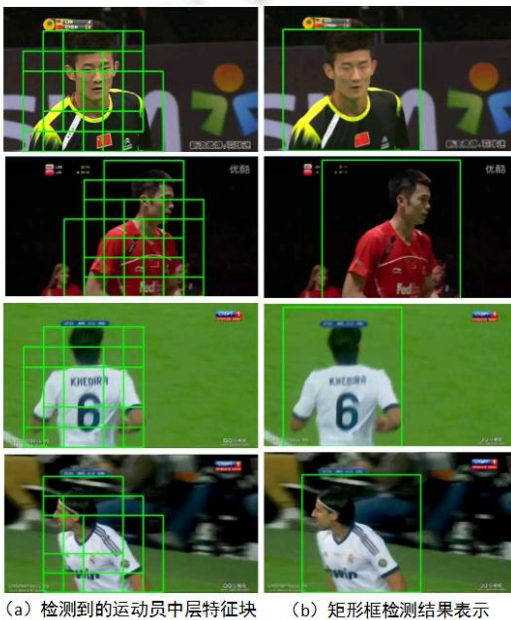


Fig.7 Player detection result of close-up shot  
图 7 特写镜头运动员检测结果

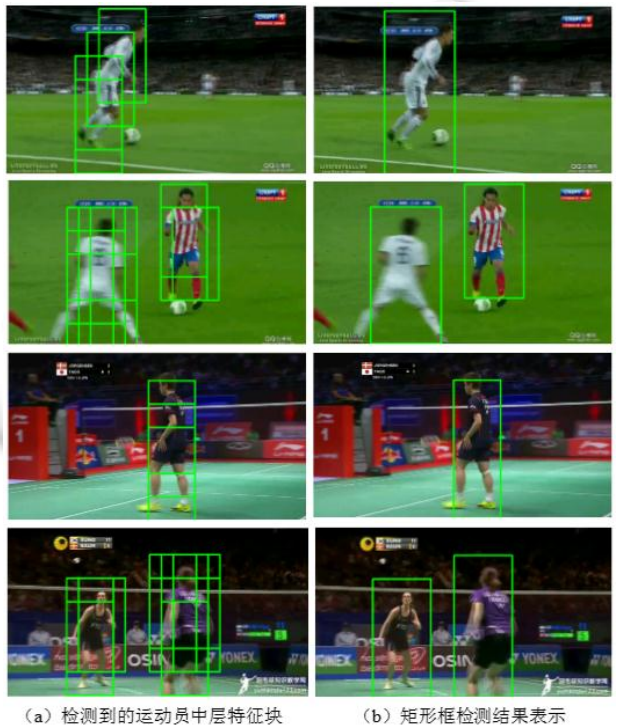


Fig.8 Player detection result of medium view shot  
图 8 中镜头运动员检测结果



### 3. 远镜头运动员检测

选取视频帧尺寸为  $512 \times 288$  的足球、乒乓球和羽毛球视频作为采样和测试的视频.远镜头中场地率最大且场地色均匀.在运动员检测之前通过场地检测去除球场色,对剩余非场地区域检测运动员,以减少检测判断次数,加快检测速度.在足球、网球和羽毛球视频中,远镜头的摄像机位置距离运动员较远,图像中的运动员较小,但运动员的动作姿势对运动员的大小仍有影响.远镜头运动员的中层特征块是接近物体级别的特征.实验设置训练和检测的块大小为  $32 \times 64$ ,步长为  $8 \times 16$ .在乒乓球视频中,远镜头的摄像机位置距离运动员相对较近,图像中运动员稍大一些.实验设置训练和检测的块大小为  $64 \times 64$ ,步长为  $16 \times 16$ .

检测结果如图 9 所示:图 9(a)是场地检测结果,图 9(b)是去除场地色的结果,图 9(c)是运动员检测结果的矩形框表示;前两行是羽毛球视频中的运动员检测结果,第 3 行是乒乓球球视频中的运动员检测结果,第 4 行是足球视频中的运动员检测结果.

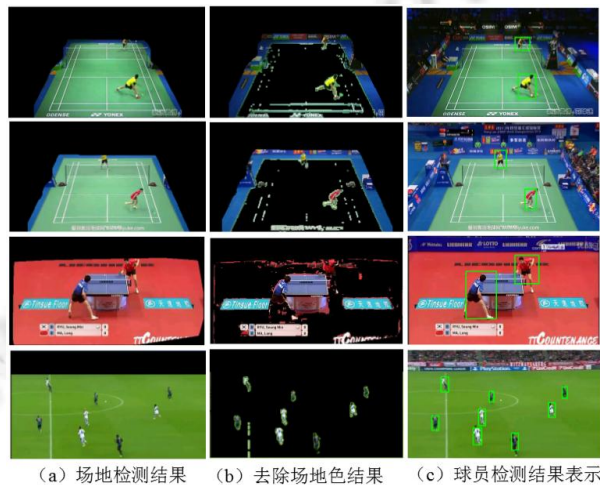


Fig.9 Player detection result of far view shot

图 9 远镜头运动员检测结果

### 4. 运动员检测结果统计与分析

表 1 是运动员检测的统计结果,由于乒乓球、羽毛球和网球的特写镜头与中镜头中的运动员检测情况类似,所以只对足球和羽毛球视频中的特写镜头和中镜头做了统计.乒乓球视频的远镜头与羽毛球、网球视频的远镜头有差别,所以实验中对足球、乒乓球和羽毛球视频做了统计.分析表 1,发现所有球类视频中特写镜头准确率都比较稳定且比较高;足球视频远镜头和羽毛球视频中镜头都有漏检的现象.其原因在于,足球视频远镜头运动员太小,或球员在球场边缘位置(如图 10 中椭圆圈起的运动员所示);羽毛球视频中镜头如图 11 所示,能够检测到矩形框中的运动员,但椭圆形中的运动员与中镜头中的一般运动员大小有较大的差异(一般运动员大小为矩形框中的运动员),造成漏检现象.

Table 1 Statistics of player detection results

表 1 运动员检测结果统计

统计项目	足球			羽毛球			乒乓球
	特写镜头	中镜头	远镜头	特写镜头	中镜头	远镜头	远镜头
总数	200	335	396	200	305	429	397
检测数	199	327	321	200	259	425	390
误检数	0	2	0	0	2	5	2
准确率(%)	98.00	97.01	81.06	100	84.26	97.90	97.73
查全率(%)	100	99.39	100	100	99.23	98.82	99.49



Fig.10 The missed detections in soccer video  
图 10 足球视频中漏检



Fig.11 The missed detections in badminton video  
图 11 羽毛球视频中镜头漏检

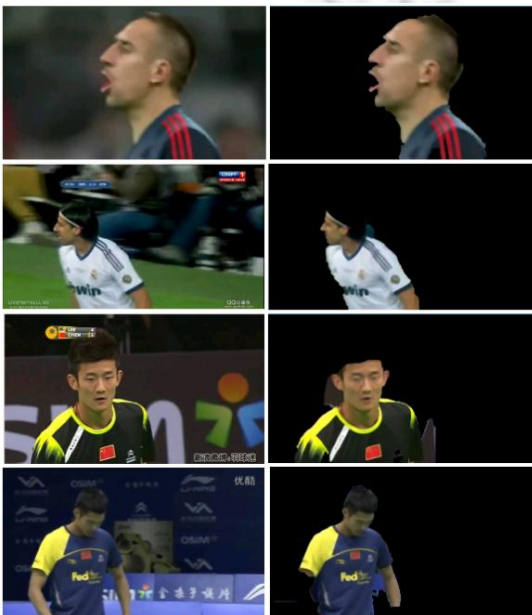
4.2 体育视频中运动员的分割

1. 特写镜头中运动员分割

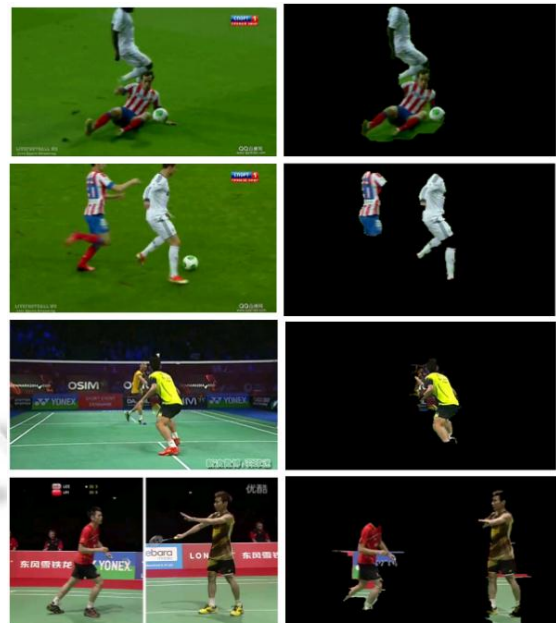
选取分辨率为 512×288 的足球和羽毛球视频作为实验视频,实验中利用运动员检测结果,设置分割超像素的个数为 60,最后使用 Grab Cut 算法分割出的结果如图 12 所示.其中,图 12(a)为原始图像,图 12(b)为分割结果;前两行为足球视频特写镜头图像的分割结果,后两行为羽毛球视频特写镜头图像的分割结果.

2. 中镜头中运动员分割

选取分辨率为 512×288 的足球和羽毛球视频作为实验视频,实验中利用运动员检测结果,设置分割超像素的个数为 150,最后使用 Grab Cut 算法分割出的结果如图 13 所示.其中,图 13(a)为原始图像,图 13(b)为分割结果;前两行为足球视频中镜头图像的分割结果,后两行为羽毛球视频中镜头图像的分割结果.



(a) 分割前图像 (b) 运动员分割结果图像



(a) 分割前图像 (b) 运动员分割结果图像

Fig.12 Player segmentation result of close-up shot  
图 12 特写镜头运动员分割结果

Fig.13 Player segmentation result of medium view shot  
图 13 中镜头运动员分割结果

3. 远镜头中运动员分割

选取分辨率为 512×288 的足球、羽毛球和乒乓球视频作为实验视频.由于足球、羽毛球视频远镜头中场地颜色比较均匀,所以实验直接采用检测结果的 mask1 与经过场地色消除后的结果 mask2 做与操作,得到运动员分割结果.在乒乓球视频中,由于运动员相对较大,所以仍然采用超像素标记.利用 Grab Cut 算法得到最终分割结果,实验中超像素个数设置为 200.分割结果如图 14 所示,其中,图 14(a)为原图像,图 14(b)为分割结果;第 1 行为



足球视频远镜头图像的分割结果,第 2 行为羽毛球视频远镜头图像的分割结果,第 3 行为乒乓球视频远镜头图像的分割结果。

#### 4. 运动员分割结果分析

各类视频的分割结果表明,运动员的大部分轮廓都能够被分割出来.不足的地方主要是运动员的头发和胳膊存在缺失现象,另外,乒乓球远镜头中存在球台残留现象.原因在于:(1) 运动员头发颜色与背景颜色很接近;(2) 由于运动员胳膊比较狭长,超像素分割时可能将胳膊上的点与边缘背景点聚类成一个超像素;(3) 头发和胳膊位于运动员的轮廓边缘,运动员区域检测结果不一定完整覆盖.在乒乓球视频远镜头中,直接利用运动员的矩形区域检测结果标记超像素造成球台残留.针对这一不足,可以利用球台区域较大且颜色均匀的特点,在比赛场地检测后通过检测球台颜色去除球台像素点,从而更准确地分割出运动员。



(a) 分割前图像 (b) 运动员分割结果图像

Fig.14 Player segmentation result of far view shot

图 14 远镜头运动员分割结果

## 5 总 结

本文分析总结了体育视频的特点,结合中层特征块的性质,提出半监督挖掘运动员的中层特征块,训练基于中层特征块的运动员检测分类器,对不同的镜头类型使用不同的分类器检测运动员.根据运动员检测结果与其轮廓之间的关系,利用超像素和 Grab Cut 分割算法实现运动员的分割.实验结果表明,基于中层特征块的运动员区域检测算法具有较高的准确率.与基于底层特征的运动员区域检测结果相比,基于中层特征块的检测结果能够更加精确地表示运动员区域.但如何更有效利用中层特征块,有待进一步研究.基于超像素的运动员轮廓分割算法既能有效地保持图像中的运动员轮廓信息,又能减少计算量.但由于部分运动员头发、胳膊和手臂等边缘部分的超像素分类不够准确,导致分割效果不够理想,需要进一步改进。

## References:

- [1] Chen FD, Hong BR. Object detection method based on background image difference using dynamic threshold. Journal of Harbin Institute of Technology, 2005,37(7):883-884 (in Chinese with English abstract).
- [2] Stauffer C, Grimson WEL. Learning patterns of activity using real time tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000,22(8):747-757.
- [3] Sun JJ, Lian YN, Tian YM. A method of object detection based on improved Gaussian mixture model. In: He X, *et al.*, eds. Computer Informatics Cybernetics and Applications. Springer-Verlag, 2012. 673-679.
- [4] Elgammal A, Harwood D, Davis L. Non-Parametric model for background subtraction. In: Proc. of the 6th IEEE European Conf. on Computer Vision. Boston, 2000. 751-767.
- [5] Cui XY, Huang JZ, Zhang ST, Metaxas DN. Background subtraction using low rank and group sparsity constraints. In: Proc. of the ECCV 2012. Berlin, Heidelberg: Springer-Verlag, 2012. 612-625.
- [6] Elqursh A, Elgammal A. Online moving camera background subtraction. In: Proc. of the ECCV 2012. Berlin, Heidelberg: Springer-Verlag, 2012. 228-241.
- [7] Sheikh Y, Javed O, Kanade T. Background subtraction for freely moving cameras. In: Proc. of the 12th IEEE Int'l Conf. on Computer Vision. 2009. 1219-1225.
- [8] Gavrilin DM, Philomin V. Real-Time object detection for smart vehicles. In: Proc. of the 7th IEEE Int'l Conf. on Computer Vision. 1999. 87-93.
- [9] Papageorgiou C, Poggio T. A trainable system for object detection. International Journal of Computer Vision, 2000,38(1):15-33.
- [10] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. San Diego, 2005. 886-893.

- [11] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained multiscale, deformable part model. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2008. 1–8.
- [12] Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D. Object detection with discriminatively trained part based models. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010,32(9):1627–1645.
- [13] Felzenszwalb PF, Girshick RB, McAllester D. Cascade object detection with deformable part models. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. San Francisco, 2010. 2241–2248.
- [14] Schick A, Bäuml M, Stiefelhagen R. Improving foreground segmentations with probabilistic superpixel Markov random fields. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. 2012. 27–31.
- [15] Fulkerson B, Vedaldi A, Soatto S. Class segmentation and object localization with superpixel neighborhoods. In: Proc. of the 12th IEEE Int'l Conf. on Computer Vision. 2009. 670–677.
- [16] Singh S, Gupta A, Efros AA. Unsupervised discovery of mid-level discriminative patches. In: Proc. of the ECCV 2012. Berlin, Heidelberg: Springer-Verlag, 2012. 73–86.
- [17] Rother C, Kolmogorov V, Blake A. Grabcut: Interactive foreground extraction using iterated graph cuts. ACM Trans. on Graphics, 2004,23(3):309–314.
- [18] Zhang YZ, Wang JY, Dai YW. Soccer video shot segmentation based on self-adapting dual threshold and dominant color percentage. Journal of Nanjing University of Science and Technology (Natural Science), 2009,33(4):432–437 (in Chinese with English abstract).
- [19] Yu JQ, Wang N. Shot classification for soccer video based on sub-window region. Journal of Image and Graphics, 2008,13(7):1347–1352 (in Chinese with English abstract).
- [20] Ren XF, Malik J. Learning a classification model for segmentation. In: Proc. of the 9th IEEE Int'l Conf. on Computer Vision. 2003. 10–17.
- [21] Shi JB, Malik J. Normalized cuts and image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000,22(8):888–905.
- [22] Moore AP, Prince SJD, Warrell J, Mohammed U, Jones G. Superpixel lattices. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. San Francisco, 2008. 1–8.
- [23] Vedaldi A, Soatto S. Quick shift and kernel methods for mode seeking. In: Proc. of the European Conf. on Computer Vision (ECCV 2008). 2008. 705–718.
- [24] Levinshtein A, Stere A, Kutulakos KN, Fleet DJ, Dickinson SJ. Turbopixels: Fast superpixels using geometric flows. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(12):2290–2297.
- [25] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. Slic superpixels. Ecole Polytechnique Federal Laussanne, 2010,30(2):149–300.

#### 附中文参考文献:

- [1] 陈凤东,洪炳镕.基于动态阈值背景差分算法的目标检测方法.哈尔滨工业大学学报,2005,37(7):883–884.
- [18] 张玉珍,王建宇,戴跃伟,等.基于自适应双阈值和主色率的足球视频镜头的分割.南京理工大学学报(自然科学版),2009,33(4):432–437.
- [19] 于俊清,王宁.基于子窗口区域的足球视频镜头分类.中国图象图形学报,2008,13(7):1347–1352.



姚沁汝(1988—),女,安徽安庆人,硕士,主要研究领域为数字视频分析与处理。



于俊清(1975—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为多媒体处理与检索,多核计算与流编译。



唐九飞(1970—),男,工程师,CCF 专业会员,主要研究领域为视频信息分析与管理。



王赠凯(1980—),男,博士,主要研究领域为多媒体处理与检索。