

基于子窗口区域的足球视频镜头分类

于俊清 王宁

(华中科技大学计算机科学与技术学院, 武汉 430074)

摘要 为了对海量视频数据进行有效的管理和快速浏览,急需对数字视频进行基于内容的视频检索。镜头分类是足球视频处理与检索的重要部分,针对目前现有足球镜头分类方法存在算法准确性不高或运算量过大的问题,提出了一种新的基于子窗口区域的镜头分类方法。该方法采用在 HSV 颜色空间中计算足球视频帧子窗口区域球场景色像素比率,并辅以边缘信息的检测,对足球视频中的主镜头、中镜头、特写镜头和其他镜头进行了分类,实验结果表明该方法切实可行,具有很高的检出率和准确率。

关键词 镜头分类 子窗口区域 足球视频 HSV 颜色空间

中图法分类号:TP391 文献标识码:A 文章编号:1006-8961(2008)07-1347-06

Shot Classification for Soccer Video Based on Sub-window Region

YU Jun-qing, WANG Ning

(College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074)

Abstract It is urgently needed to develop a solution to content-based video retrieval, in order to support efficient management and quick browsing mass video data. Shot classification is an important research content in soccer video processing and retrieval. Considering the shortcomings of the existed methods, we propose a novel approach which can classify soccer video shots into main, middle, close-up shots and other ones based on sub-window region. The key step is to calculate the proportion of pixels with field color in sub windows region of soccer video frame in HSV color space. Experimental results have shown that the proposed method has a very high precision and recall rate.

Keywords shot classification, sub-window region, soccer video, HSV color space

1 引言

足球作为世界第一大球类运动深受人们的喜爱,对于足球视频的管理与检索也成为体育视频研究的一个重要内容。由于足球视频的特殊性,足球视频的处理与一般的新闻视频和电影视频的处理方法不同。足球视频的研究主要包括:镜头边界检测、镜头分类、慢镜头检测、提取精彩镜头、球场场景重建、事件检测和视频摘要等。

足球视频镜头分类是事件检测、视频摘要生成的基础,是视频检索的必要条件之一,因此研

究视频镜头分类具有重要的理论研究和实际应用意义。视频摘要是视频数据的内容总结,在表现形式上,它是一个静止图片序列或视频片段序列^[1]。视频摘要可以通过选取视频中有代表性的图片序列或代表镜头来产生。由于视频数据量巨大,从每一个镜头中选取关键帧或选取镜头的一部分作为视频摘要,必然存在着大量的冗余内容,尤其是在足球视频中,视频镜头的内容在播放过程中存在着较大的重复性,因此有必要根据镜头的内容进行分类,在每一类镜头中选取一定比例的关键镜头或关键帧作为视频摘要。足球视频摘要生成过程包括镜头边界检测、慢镜头

基金项目:国家自然科学基金项目(60703049);中国博士后基金项目(20060400847);湖北省自然科学基金项目(2007ABA265)

收稿日期:2007-08-15;改回日期:2007-12-05

第一作者简介:于俊清(1975~),男,副教授。现为华中科技大学计算机科学技术专业在站博士后。研究方向为基于内容的多媒体信息处理与检索。E-mail:yjqing@hust.edu.cn

检测、镜头分类和事件检测 4 个步骤^[2],详细流程如图 1 所示。由该图可以看出,足球视频镜头分类在足球视频摘要生成的过程具有举足轻重的地位。

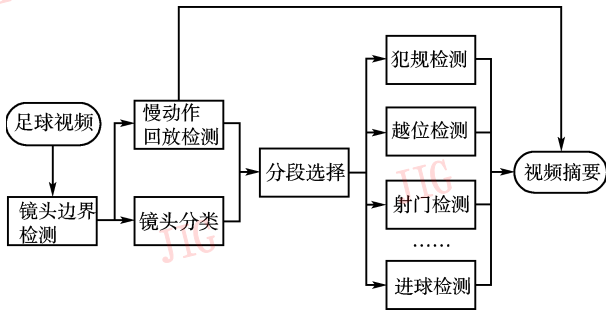


图 1 足球视频摘要处理流程

Fig. 1 Processing diagram of soccer video abstract

2 足球视频镜头分类

2.1 常见的镜头类型

根据视频图像表现出的视域范围可以将足球镜头分为 4 种类型:主镜头类型(Main type)、中镜头类型(Middle type)、特写镜头类型(Close-up type)和其他镜头类型(Other type)。

主镜头类型 由主摄像机拍摄的足球比赛远镜头,球场区域所占面积较大,人物目标很小,如图 2(a)、(b)所示。

中镜头类型 由场边长焦镜头摄像机拍摄的以球员为目标的镜头,人物目标为全身像,如图 2(c)、(d)所示。

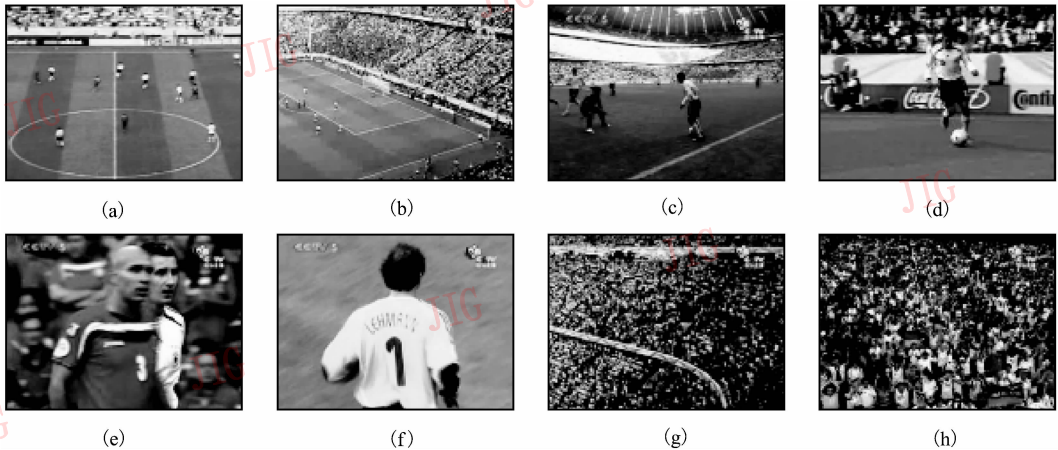


图 2 足球视频镜头类型示例

Fig. 2 Illustrations of soccer video shots

特写镜头类型 由场边长焦镜头摄像机拍摄的镜头,以球员、裁判、教练或观众为目标,人物目标为半身像,如图 2(e)、(f)所示。

其他镜头类型 一般以观众为目标的远镜头,人物目标颗粒度很小,如图 2(g)、(h)所示。

2.2 现有的主要分类方法

由于镜头分类在足球视频分析处理中具有非常重要的意义,因此受到了众多研究者的关注,也已经提出了很多不同的分类方法。文献[3]中提出采用球场颜色比率为特征的足球视频镜头分类方案。该方案根据球场颜色占整个屏幕的比率设定一个阈值,将镜头分为长距镜头、中距镜头和其他镜头。该方法对于中距镜头与长距镜头颜色比例相似的情况,分类效果不理想。同时由于该方

法对整幅帧图像进行处理,其运算量较大,难以实现实时处理。

文献[4]中提出了一种利用黄金分割原理进行足球视频镜头分类的方法。该方法将一幅帧图像分割为大小不等的 9 个区域,根据不同区域的颜色差别利用贝叶斯分类器对长距镜头、中距镜头和特写与场外镜头进行分类。与本文方法相比,主要缺点是对视频帧不同区域分别计算颜色差,会带来较大的运算量。

文献[5]中提出的足球视频镜头分类方法与文献[3]中的方法类似,也是计算整个帧图像球场颜色所占比率,设定一组阈值,对远镜头、中镜头和特写与场外镜头进行分类。为了能够更准确地区分中镜头和远镜头,采用了寻找非球场色最大矩形区域

方法。该方法在区分中镜头和远镜头时,确定非球场合最大矩形区域的计算量很大,难以对视频进行实时处理。

文献[6]、[7]中分别利用支持向量机(SVM)和隐马尔可夫模型(HMM)对足球体育视频进行语义分类。采用 SVM 和 HMM 进行镜头分类,虽然在准确率上有一定的提高,但算法实现起来较为复杂,计算量较大。

3 基于子窗口区域的分类方法

在系统分析现有方法各自特点的基础上,通过仔细观察足球视频各类镜头的典型特征,提出基于子窗口区域的足球视频镜头分类方法。该分类方法利用子窗口区域的球场颜色比率来反映镜头的类型。一般来说,主镜头和以球场为背景的中镜头类型的球场颜色比率较高,而特写镜头和观众镜头的球场颜色比率比较低。实验结果表明,不同镜头类型的球场颜色比率的变化在图 3 中所示子窗口 2 内反映的更加明显。所以采用子窗口矩形区域进行球场颜色计算,既能降低运算复杂度,又能保留整幅帧图像的颜色分布特征。图 3 中子窗口 1 不仅用于区分主镜头类型和非主镜头类型,还用于区分特写镜头类型和其他镜头类型;子窗口 2 用于把被认为是非主镜头类型的主镜头从非主镜头类型划分出去;子窗口 3 用于在非主镜头类型中区分中镜头类型和非中镜头类型。

3.1 不同镜头类型的特点和区分方法

主镜头类型与非主镜头类型有一些不同之处:主镜头帧下半部分像素球场合所占比率较大,其中球员、足球和裁判等非球场合像素较为分散且所占比率较小。因此可以确定一个合适的子窗口区域,除去观众席、球门和教练席等非场地区域的影响,使得所有主镜头类型在这个区域内的都是球场合。在计算这个子窗口区域内的球场合像素比率时,主镜头类型与非主镜头类型会有明显差异。实验结果证明,选择图 3 所示子窗口 1 作为统计区域,会取得较好的效果。如果仅采用子窗口 1 和阈值 T_1 ,会使为球场合的中镜头的少部分被认为是主镜头类型,产生误检。所以在这里采用了双阈值方法:采用 T_1 和 T_1' ,利用子窗口 2 进行修正。

在非主镜头类型中区分中镜头和非中镜头时,要考虑到特写镜头其自身的特点。相对于特写镜头类型和其他镜头类型,中镜头帧最下端区域球场合所占比例较大。实验结果表明,采用图 3 所示子窗口 3 作为区分中镜头类型和非中镜头类型的窗口区域效果非常好。

对于其他类型的镜头(比如观众镜头)与主镜头类型、中距镜头类型和特写镜头类型不同。由于其目标颗粒度较小,所以同等尺寸窗口中具有更多的边缘信息。在检测其他镜头类型时,对窗口区域采用 Canny 算子边缘二值化统计边缘像素比率。在本方案中,采用了图 3 所示子窗口 1 区域。

3.2 子窗口球场颜色统计方法

统计子窗口区域的球场颜色比率是在 HSV 色彩空间中完成的。在 HSV 色彩空间中,每个颜色值由一个色调(hue)、饱和度(saturation)和亮度(value)的构成的三元组(h, s, v)来表示。

HSV 空间具有两个重要特点:一是人眼可独立感知该空间各颜色分量的变化;二是在该颜色空间中的颜色三元组之间的欧几里德距离与人眼感觉到的相应的颜色差具有线性关系,是一种符合人类视觉感知特性的颜色模型。因此基于 HSV 色彩空间的测度能够更好地逼近人眼的感觉,这是选择 HSV 色彩空间作为提取镜头颜色特征的重要原因。另外,对于足球场来说,整个场地基本上是由球场合像素构成,利用色调分量即可统计球场合像素,从而在保证准确性的同时提高了算法的效率。在本方法中,统计球场合像素比率时,色调分量 H 范围设定在 $75^\circ \sim 105^\circ$ 之间,球场合在 HSV 颜色空间中的位

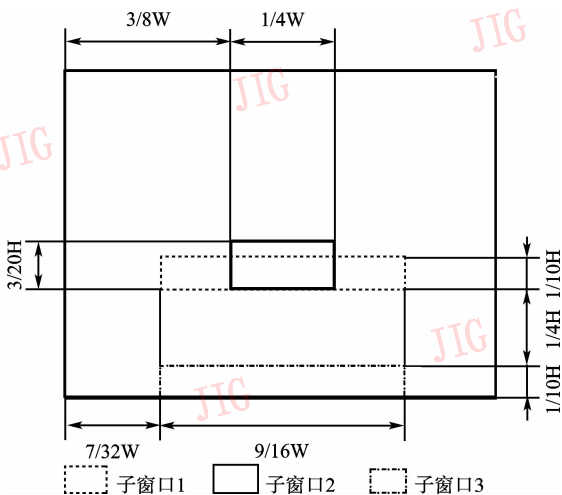


图 3 子窗口在帧图像中的位置和大小

Fig. 3 Position and dimension of sub-windows in frame

置如图 4 所示。

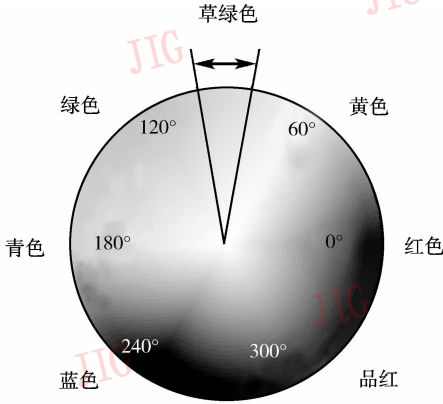


图 4 球场色在 HSV 颜色空间中的位置

Fig. 4 Distribution of field color in HSV color space

3.3 子窗口区域尺寸定义

在进行镜头分类时,不是直接对整幅帧图像进行计算,而是采用图 3 所示 3 个子窗口区域进行镜头分类。这 3 个子窗口区域的位置和大小是经过大

量足球视频分类实验而确定的,子窗口区域的选择可以大大简化像素的计算量。

在图 3 中,对于子窗口 1,主镜头在这个区域内不包括观众席与球门,且在其它类型镜头中,这个区域全是观众像素。对于子窗口 2,中镜头在这个区域内比特写镜头包括更多的球场像素。主镜头在子窗口 3 内不包括观众席与球门像素,所以子窗口 3 的作用是修正子窗口 1 带来的误检。

3.4 算法流程

算法流程如图 5 所示。

输入:以已经检测了镜头边界的足球视频和该视频的镜头标记文件作为输入数据,即输入为镜头集 S 。

$$S = \{shot_i\} \quad i = 0, 1, 2, \dots, N - 1 \quad (1)$$

其中,每一个镜头 $shot_i$ 包含起始帧号 f_{start} 、终止帧号 f_{end} 和关键帧号 f_{key} 。

$$f_{key} = \lceil (f_{start} + f_{end}) / 2 \rceil \quad (2)$$

输出:将输入的镜头作镜头类型标记。

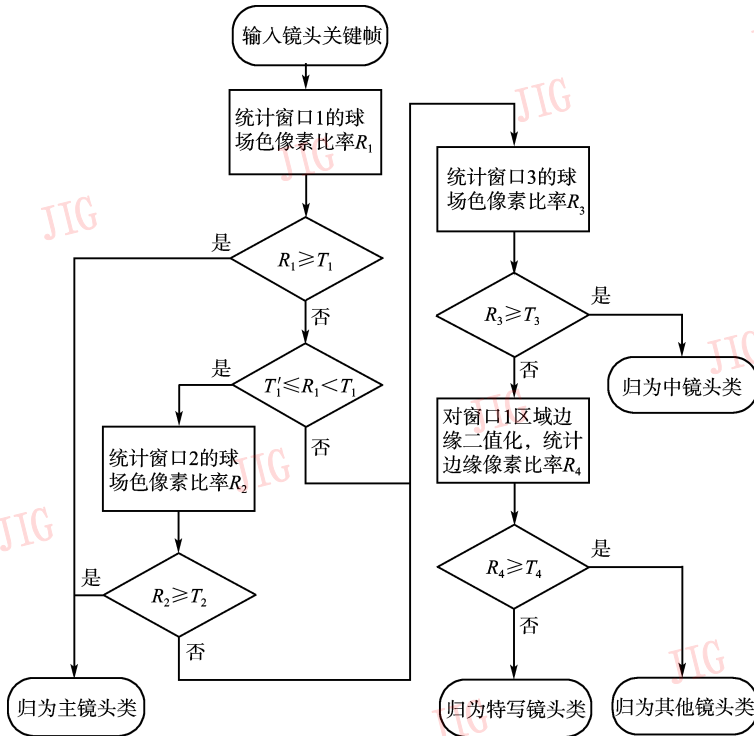


图 5 镜头分类算法流程图

Fig. 5 Diagram of shot classification algorithm

具体步骤如下:

设定循环变量 i 从 0 到 $N - 1$, N 为镜头数。

(1) 取得帧号为 f_{key} 的帧图像 P_i ;

(2) 对图像 P_i 按图 3 中所示窗口区域 1 取得像

素数据,由 RGB 空间转化到 HSV 空间,统计其球场色所占比率 R_1 ;

(3) 如果 $R_1 \geq T_1$,则认为 $shot_i$ 为主镜头类型转第 4 步⑦,否则继续;

(4) 如果 $T'_1 \leq R_1 < T_1$, 则进行如下处理:

① 对图像 P_i 按图 3 中所示子窗口 2 取得像素数据, 由 RGB 空间转化到 HSV 空间, 统计其球场色所占比率 R_2 ;

② 如果 $R_2 \geq T_2$, 则认为 $shot_i$ 为主镜头类型, 转⑦, 否则继续;

③ 图像 P_i 按图 3 中所示子窗口 3 取得像素数据, 由 RGB 空间转化到 HSV 空间, 统计其球场色所占比率 R_3 ;

④ 如果 $R_3 \geq T_3$, 则认为 $shot_i$ 为中镜头类型, 转⑦, 否则继续;

⑤ 对图像 P_i 按图 3 中所示子窗口 1 取得像素

数据, 利用 Canny 算子边缘二值化处理, 统计其边缘点像素所占比率 R_4 ;

⑥ 如果 $R_4 \geq T_4$, 则认为 $shot_i$ 为其他镜头类型, 否则认为 $shot_i$ 为特写镜头类型;

⑦ 如果 $i = N - 1$, 则结束, 否则令 $i = i + 1$, 转到①。

3.5 阈值的确定

实验中对足球镜头进行处理, 将各子窗口球场像素比率分布和边缘像素比率分布进行统计, 实验结果如图 6 所示。通过实验可以很好地确定镜头分类阈值。取 $T_1 = 0.95$, $T'_1 = 0.8$, $T_2 = 0.9$, $T_3 = 0.7$ 和 $T_4 = 0.25$ 。

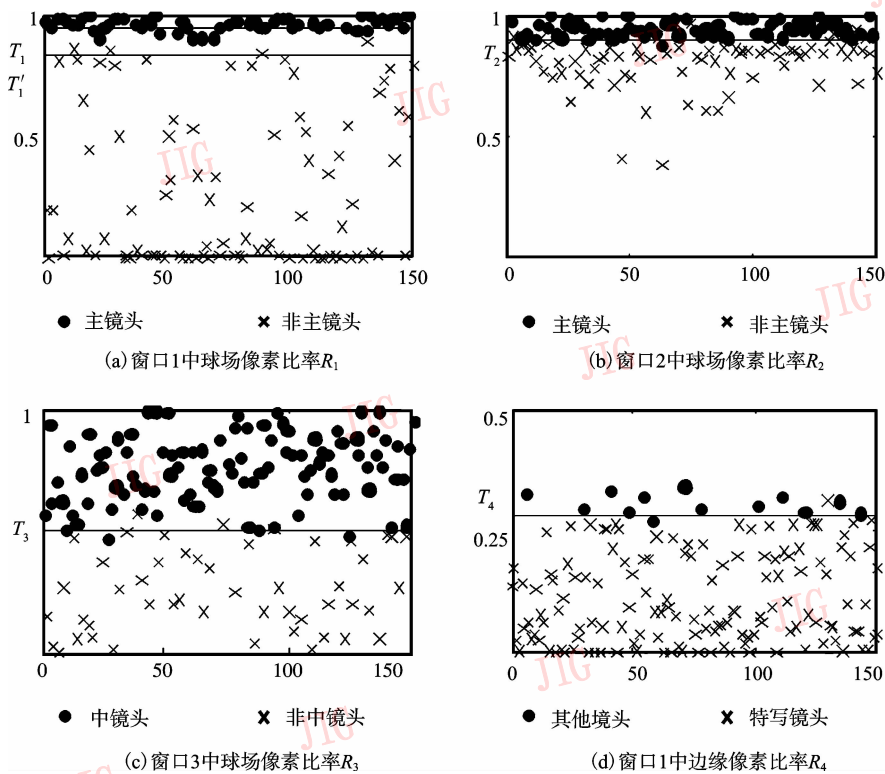


图 6 镜头分布与阈值确定示意图

Fig. 6 Illustration of Shot distribution and threshold setting

4 实验结果

为了验证算法的有效性, 实验中选择了多段不同类型的足球比赛视频, 其中包含有 1 000 余个足球视频镜头。实验结果如表 1 所示。从表 1 中可以看出该分类方法对于主镜头、中镜头、特写镜头和其他类型镜头都得到了较好的实验结果。由于中镜头类型人物目标的动作过大, 有时会导致图 3 中子窗口 1 无

法捕捉目标。这样使得中镜头类型被误判为主镜头类型, 在一定程度上影响了分类结果的正确率。

表 1 足球视频镜头分类结果

Tab. 1 Experimental results of shot classification

类型	镜头数	检测数	正确数	正确率 (%)	查全率 (%)
主镜头	406	385	357	92.7	87.9
中镜头	308	315	287	91.1	93.2
特写镜头	105	119	98	82.3	93.3
其他类镜头	21	21	21	100	100

5 结 论

利用子窗口区域代替整幅帧图像来对足球视频分类,不仅分类效果很好,而且大大降低计算复杂度,提高了算法的计算效率,可以达到实时处理。因此该方法可以应用在足球视频播放、处理、检索等实时系统中。

值得注意的是,区分其他镜头类型和特写镜头时,除了通过计算窗口区域的边缘像素比率的方法外,还可以利用纹理特征方法实现。不同镜头类型子窗口区域中图案纹理特征不同,比如观众镜头中图案纹理细密、颗粒度小,所以利用纹理特征也可以达到很好的效果。在 MPEG-7 标准中的纹理浏览特征可以更精确地反映图案纹理和颗粒度信息,可以用来取代利用 Canny 算子统计边缘像素方法,具体实现可以参考文献[8]、[9]。

参考文献 (References)

- 1 Zhou Yi-hua, Cao Yuan-da, LI Jian, *et al.* Soccer video shot classification method based on color and edge distribution [J]. Transactions of Beijing Institute of Technology, 2005, **25** (12): 1079 ~ 1082. [周艺华, 曹元大, 李剑等. 基于颜色与边缘分布的足球视频镜头分类方法[J]. 北京理工大学学报, 2005, **25**(12):1079 ~ 1082.]
- 2 Ahmet Ekin, Tekalp A Murat, Rajiv Mehrotra. Automatic soccer video analysis and summarization [J]. IEEE Transactions on Image Processing, 2003, **12**(7):796 ~ 807.
- 3 Xu Peng, Xie Le-xing, Chang Shih-Ful. Algorithms and system for segmentation and structure analysis in soccer video [A]. In: Proceedings of IEEE International Conference on Multimedia and Expo [C], Tokyo, Japan, 2001: 721 ~ 724.
- 4 Ahmet Ekin, Tekalp A Murat. Shot type classification by dominant color for sports video segmentation and summarization [A]. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing [C], Hong kong, China, 2003:173 ~ 176.
- 5 Zhao Pi-xi, Hu Bin, Wang Xiu-kun, *et al.* Analysis and summarization for soccer video [J]. Journal of Computer Engineering & Application, 2005, **41**(30): 166 ~ 168. [赵丕锡, 胡滨, 王秀坤等. 足球视频的结构分析与概要[J]. 计算机工程与应用, 2005, **41**(30): 166 ~ 168.]
- 6 Zhou Yi-hua, Cao Yuan-da, Zhang Long-fei, *et al.* An svm-based soccer video shot classification [A]. In: Proceedings of International Conference on Machine Learning and Cybernetics [C], Guangzhou, China, 2005: 5398 ~ 5403.
- 7 Xavier Gibert, Li Hui-ping, David Doermann. Sports video classification using hmms [A]. In: Proceedings of International Conference on Multimedia and Expo [C], Baltimore, Maryland, 2003, **2**:245 ~ 248.
- 8 ISO/IEC 15938-3. Multimedia Content Description Interface Part 3: Visual [S].
- 9 ISO/IEC 15938-8. Multimedia Content Description Interface Part 8: Extraction and use of MPEG-7 descriptions [S].