

自适应高斯混合模型球场检测算法及其在体育视频分析中的应用

刘 扬¹ 黄庆明² 高 文^{1,3} 叶齐祥²

¹(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

²(中国科学院研究生院 北京 100080)

³(北京大学信息科学技术学院 北京 100871)

(liuy@vilab.hit.edu.cn)

Playfield Detection Using Adaptive GMM and Its Application in Sports Video Analysis

Liu Yang¹, Huang Qingming², Gao Wen^{1,3}, and Ye Qixiang²

¹(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

²(Graduate University, Chinese Academy of Sciences, Beijing 100080)

³(School of Electronics Engineering and Computer Science, Peking University, Beijing 100871)

Abstract Playfield detection is a key step in sports video content analysis, while the playfield presents different colors in broadcast video as a result of illumination variation caused by different cameras and different shooting angles, etc. To address the problem, adaptive GMM is exploited to detect playfield. The proposed algorithm first selects samples in a video clip randomly and extracts the dominant color in the histogram automatically; then the dominant color is modeled by GMM approximately. To adapt the model to the playfield color variation, the model's parameters are updated by incremental EM algorithm in the process of playfield detection. Based on the playfield's detection result, a simple and effective algorithm is proposed to classify the scene in soccer video according to the playfield distribution region in a frame. Experimental results show that the proposed algorithms are effective.

Key words playfield detection; adaptive GMM; incremental EM; soccer video; scene classification

摘 要 球场检测在体育视频内容分析中有着重要作用。为了克服由于不同光照、不同相机、不同拍摄角度造成球场颜色的非均匀性问题,提出了一种基于自适应高斯混合模型(adaptive Gaussian mixture model, GMM)的球场检测算法。该算法首先从视频中任意抽取一些图像,并自动分析这些图像的主要颜色,从中找到主颜色的近似分布,然后,利用 GMM 拟合主要颜色分布。为提高模型的适应能力,在球场检测过程中,利用当前 GMM 球场检测结果和增量期望最大(incremental expectation maximum, IEM)算法不断更新模型参数,从而得到更加准确的参数估计,并用于后续图像中球场和非球场像素进行分类。最后,根据球场区域在图像中的分布,对足球比赛场景进行分类。实验证明,提出的算法具有良好的性能。

关键词 球场检测;自适应高斯混合模型;增量期望最大算法;足球视频;场景分类

中图法分类号 TP391.41

1 引 言

体育节目拥有大量的观众。为了能够快捷地访

问含有精彩场面或重要事件的体育视频片断,近年来许多研究人员投入到体育视频内容分析这个研究领域。球场检测对体育视频内容分析具有非常重要的作用。例如,体育视频场景分类、对象(包括球、球员

和标线等)检测,以及运动视频结构分析等都非常依赖于球场检测结果.因此,球场检测引起了研究人员的重视.

Luo 等人^[1]利用事先设定的绿色球场在 HSI 空间分布区域来检测球场,这种方法虽然能够较好地检测到球场区域,但也有很大的局限性.因为不同的体育比赛球场颜色是不同的,即便都是足球比赛,各个球场的颜色也是不尽相同的. Ekin 等人^[2]首次对球场检测的算法进行了比较详细的介绍.在其工作中,作者采用两个颜色空间(可以从 HSI, RGB, $YCbCr$ 等颜色空间选择),一个作为基本颜色空间,另一个作为控制空间,起到对前者控制和补充的作用.在基本颜色空间中,利用球场的颜色均值和阈值来检测球场像素.作为本文的前期工作, Jiang^[3]和 Wang^[4]等人采用了高斯混合模型检测球场区域,但它们都无法自适应调整模型参数.

球场检测在体育视频分析中起到了举足轻重的作用. Xie 等人^[5,6]利用主颜色信息分析足球视频结构. Ekin 等人^[7]利用球场检测结果对体育视频中的镜头进行分类.文献[8~10]通过利用场地检测结果,进一步检测视频图像中的运动目标,包括球员和球. Assfalg 等人^[11]则在球场检测基础上,对足球视频进行了语义层次分析,自动提取精彩片断.

本文采用了自适应高斯混合模型来对球场颜色进行建模,采用此方法是基于如下考虑:

1) 一块球场上可能包含不止一种颜色.例如足球球场上可能具有条纹图案,还有一些球场上的球门区域附近由于频繁踩踏已经呈现出土地颜色.

2) 同一块球场在不同相机、不同光照、不同拍摄角度中呈现的颜色不一致.而在由多相机拍摄数

据合成的广播视频中这种现象经常遇到,因此需要模型参数能够在运行过程中适当调整.

综合以上因素,我们提出了用高斯混合模型(GMM)对球场颜色进行建模,并利用增量期望最大(IEM)算法自适应地调整参数.本方法的主要特点是:①GMM 可以更准确地描述球场颜色,尤其是当球场上颜色不一致,或存在大量噪声时,此方法的概率描述能力更强.②由于 IEM 算法的采用,能够使模型不断地适应视频中球场颜色的变化.③IEM 是一种在线学习方法,系统无需保留数量巨大的样本点.

本文将球场检测算法应用到足球视频分析当中,利用球场检测的结果,将视频中的镜头分为场内场景和场外场景.对于场内场景,本文提出通过对图像中球场区域上下边缘点的拟合,进一步将此类场景分为特写和中远景场面,并利用上边缘拟合直线的斜率检测球门附近区域.

本文所阐述的算法,是“体育视频摘要和内容丰富系统”项目(sports video summarization and enrichment system, SPISES)中的基本技术.更多的信息可以从以下网址得到:<http://www.jdl.ac.cn/en/project/mrhomepage/index.htm>.

2 基于自适应高斯混合模型的球场检测

本节首先介绍了基于自适应高斯混合模型(adaptive Gaussian mixture model, AGMM)的球场检测算法的系统框架,然后详细介绍了框架中的各个部分.

2.1 系统框架

图 1 给出了基于 AGMM 的球场检测算法框架.

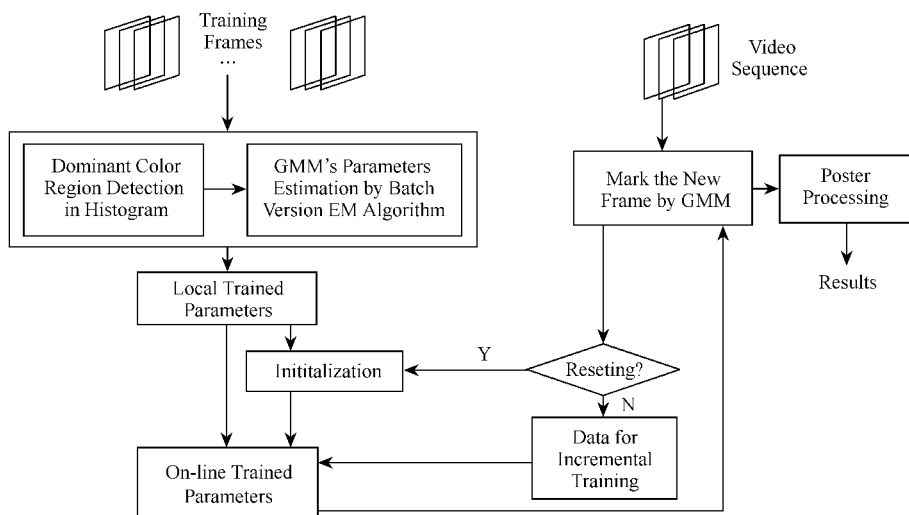


Fig. 1 The framework of the proposed playfield detection system.

图 1 球场检测系统框架图

其流程为:首先通过检测直方图中的主颜色区域,初始化 GMM 的参数,然后利用后处理得到场地区域像素,通过增量 EM 算法来修正 GMM 的参数.以下各小节详细介绍系统流程中每一个步骤.

2.2 初始样本选择

通过观察可以发现,在有球场的体育比赛节目中,一段视频图像的颜色主要分布在颜色空间中的几个比较密集区域.基于这一特点,本算法在颜色直方图中寻找一个数值普遍较大的密集区域,并将其作为球场颜色的大致分布区域.

算法抽取视频序列的前几分钟图片颜色数据,并计算所有这些像素在 $C_b C_r$ 颜色空间中的颜色直方图 H_1 . 选择 $C_b C_r$ 颜色空间基于如下两点考虑:①当前流行的视频压缩标准(如 MPEG 和 H.26L)都采用该颜色空间,这样就可以减少颜色空间转换带来的计算资源消耗;②去掉了亮度信息的 $C_b C_r$ 能够较好地描述球场区域的颜色特性,减小阴影、灯光等对亮度敏感的因素对检测结果的影响.

算法 1 描述了在 $C_b C_r$ 空间中检测主要颜色区域的过程.

需要指出的是,本算法在 $C_b C_r$ 颜色空间直方图中寻找了两次峰值附近的连通区域.这样做的目的是为了“避免‘直方图中某一颜色的数值很大,但与其连通的区域所覆盖的像素数目很少’的情况.因为在体育视频中,球场像素必然具有很大的比例,像素数目很少的区域内的像素不可能是球场像素,所以舍弃 Sum_1 和 Sum_2 中的较小者所对应的区域(这种现象通常是视频编解码过程造成的).

算法 1. 颜色空间中主颜色区域检测

1) 在 H_1 中确定最大峰值点位置 P_1 ;

2) 寻找与 P_1 四连通的点(bin)且其值大于 $s \times \text{Value}(P_1)$. 计算连通区域内所有点(bin)的值的总和 Sum_1 , 从原直方图中去掉此连通区域,余下部分记为 H_2 . 其中 s 是比例系数,本文中设为 0.05.

3) 在 H_2 中进行步骤 1)和步骤 2)操作,令 H_2 中与最大峰值点 P_2 连通的区域的点(bin)的值的总和为 Sum_2 .

4) 返回 Sum_1 和 Sum_2 中的较大者所对应的连通区域,作为球场颜色大致分布区域.

算法 1 的第 4)步返回的区域被认为是球场颜色分布的大体区域,但这一区域通常是不准确的,而且有可能有很多球场像素颜色没有被包括进来.基于这一考虑,下一小节提出用 IEM 来在线学习模型参数.

2.3 模型参数的自适应调整

如第 1 节所述,球场颜色可能不均一(如足球场

上的条纹草坪),并含有噪声;另外,由于拍摄角度、光照以及不同相机拍摄的影像,广播节目中的球场颜色会不断变化,这就要求算法采用的颜色模型应该能够描述复杂的概率分布,同时能够随着视频中球场颜色的变化不断调整其参数.采用 IEM 算法更新参数的自适应高斯混合模型具有这样的特点.

1) 高斯混合模型

高斯混合模型 p 是由多个高斯函数 p_1, p_2, \dots, p_k 线性组合而成.如式(1)所示:

$$p = \sum_{i=1}^k \pi_i p_i \quad \text{其中}$$

$$p_i(\mathbf{X}; \theta_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \cdot \exp\left(-\frac{1}{2}(\mathbf{X} - \mu_i)^T (\Sigma_i)^{-1} (\mathbf{X} - \mu_i)\right)$$

$$\text{and } \sum_{i=1}^k \pi_i = 1.$$

每个高斯函数由参数集 θ_i 决定,包括均值向量 μ_i 和协方差矩阵 Σ_i . d 是样本点 \mathbf{X} 的维数. π_i 是每个高斯函数的权重. k 是高斯项个数.集合 $\Theta = \{\pi_i, \theta_i | i=1, \dots, k\}$ 包含了所有的未知参数,它们属于某个参数空间.通常,人们用批处理的 EM 算法(EM of batch version, BEM)估计这一组参数 Θ . 本文采用了有 3 个高斯项的混合模型,其中两个用来拟合球场颜色,另一个高斯项用来拟合球场上的噪声(例如,球场草皮不好时的土地颜色).

2) 自适应高斯混合模型

为了更准确地估计视频流中球场的颜色分布,本文使用 IEM 算法在线调整 GMM 的参数.如文献[12]指出的那样,IEM 更适合处理样本点是陆续到达的情况,这正是视频分析中经常遇到的情况.以统计学的观点,EM 算法中的 E 和 M 步,可以看成是共同最大化某一个函数,这个函数如式(2)所描述,IEM 算法和 BEM 算法都可以从这个函数推导出来.

$$F(\tilde{P}, \Theta) = E_{\tilde{P}}[\ln P(y, x | \Theta)] + H(\tilde{P}), \quad (2)$$

其中 $H(\tilde{P}) = -E_{\tilde{P}}[\ln \tilde{P}(y)]$ 是分布 \tilde{P} 的熵, $E_{\tilde{P}}[\cdot]$ 是在分布 \tilde{P} 下的期望.式(2)中 y 是隐含随机变量 Y 的取值,在高斯混合模型中对应于各个正态分布的权重. x 是可观测随机变量 X 的取值,对应于像素值. Θ 是模型参数.

对于 BEM,它的 E 步和 M 步可以用方程(2)表达成如下形式,其中 t 为时间步.

$$\begin{cases} \text{E 步: 设置 } \tilde{P}^{(t)} \text{ 为 } \tilde{P} \text{ 并使 } F(\tilde{P}, \Theta^{(t-1)}) \text{ 最大;} \\ \text{M 步: 设置 } \Theta^{(t)} \text{ 为 } \Theta \text{ 并使 } F(\tilde{P}^{(t)}, \Theta) \text{ 最大.} \end{cases}$$

给定一些独立的观测样本 $x=(x_1, x_2, \dots, x_n)$, 以及它们的隐含信息 $y=(y_1, y_2, \dots, y_n)$, 函数 F 可以写成如下形式:

$$F(\tilde{P}, \Theta) = \sum_i F_i(\tilde{P}_i, \Theta),$$

其中 $F_i(\tilde{P}_i, \Theta) = E_{\tilde{P}_i}[\ln P(y_i, x_i | \Theta)] + H(\tilde{P}_i)$. (3)

IEM 即采用了式(3)这个结构. 即当每个样本执行一次 E 步, 就执行一次 M 步, 而不必等所有的数据都参与 E 步计算以后, 再执行 M 步. 也就是说, 每个样本对分布 \tilde{P} 带来的影响, 都会很快的在 M 步体现, 从而实现了在线更新参数. 如果 X 和 Y 的联合分布是指数形式的, 那么在 E 步中可以计算一个充分统计量, M 步则计算在这个统计量下 Θ 的最大似然. 令在 IEM 中用 $s_i(y_i, x_i)$ 来表示第 i 个样本点的统计量, 其 E 步和 M 步可以表达如下:

E 步: 选择第 i 个样本点,

使 $\tilde{s}_j^t = \tilde{s}_j^{t-1}$, 如果 $j \neq i$,

使 $\tilde{s}_i^t = E_{\tilde{P}_i}[s_i(y_i, x_i)]$, 其中 $\tilde{P}_i(y_i) = P(y_i | x_i, \Theta^{t-1})$,

使 $\tilde{s}^t = \tilde{s}^{t-1} - \tilde{s}_i^{t-1} + \tilde{s}_i^t$.

M 步: 使 Θ^t 为 Θ , Θ 为 S^t 下的最大似然.

依据 IEM, 自适应高斯混合模的迭代过程如下:

$$\begin{aligned} \hat{\pi}_k^{N+1} &= \hat{\pi}_k^N + \frac{1}{N+1}(\hat{p}(w_k | x_{N+1}) - \hat{\pi}_k^N), \\ \hat{u}_k^{N+1} &= \hat{u}_k^N + \frac{\hat{p}(w_k | x_{N+1})}{\sum_{i=1}^{N+1} \hat{p}(w_k | x_i)}(x_{N+1} - \hat{u}_k^N), \\ \hat{\Sigma}_k^{N+1} &= \hat{\Sigma}_k^N + \frac{\hat{p}(w_k | x_{N+1})}{\sum_{i=1}^{N+1} \hat{p}(w_k | x_i)}((x_{N+1} - \hat{u}_k^N)(x_{N+1} - \hat{u}_k^N)^T - \hat{\Sigma}_k^N), \end{aligned} \quad (4)$$

其中 $k=1, 2, 3$, $\hat{p}(w_k | x_i) = \frac{\hat{p}(x_i; \theta_k)}{\hat{p}(x_i)}$.

从式(4)可以看到, 每个新样本的加入都会对模型参数产生影响, 从而使模型在运行中能够不断地逼近真实分布. 算法本身没有过多增加运算量, 因为在像素标记过程中, $\hat{p}(x_i; \theta_k)$ 和 $\hat{p}(x_i)$ 都已经计算了, 算法无需重新计算. 式 $\sum_{i=1}^{N+1} \hat{p}(w_k | x_i)$ 则可以被记录下来, 并不断累加. 增加的计算只是 $(x_{N+1} - \hat{u}_k^N)(x_{N+1} - \hat{u}_k^N)^T$, 对于 2 维的 $C_b C_r$ 像素值, 有 4 次乘法运算.

2.4 增量训练数据的选择

根据式(4), 在线更新模型参数的关键问题是样本的选择. 首先利用当前参数 Θ_c 将图像像素二值化为球场和非球场像素, 并利用形态学滤波^[13] (用 3×3 的模板对二值图像进行开运算和闭运算) 得到球场区域. 球场区域的像素将作为增量训练数据, 参加模型参数更新运算, 从而使模型能够逐渐学习到球场上的所有像素在不同相机、不同视角和不同光照下的颜色分布. 图 2 给出了示意图.

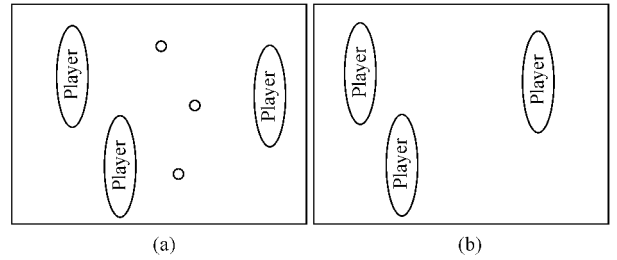


Fig. 2 Incremental training data. (a) The circles are noise regions and (b) is the result after morphological processing, and the pixels in the circles are used as incremental training data.

图 2 增量训练集. (a) 圆圈表示噪声, 经形态学滤波处理后它们被去除. (b) 处理后的图像, 圆圈内的像素作为增量训练集合

2.5 对参数漂移的处理

与所有的自适应系统一样, 在参数调整过程中可能会造成估计得到的参数向远离真实值的方向漂移. 本算法中采用如下机制判断是否产生漂移, 每隔 30s 用初始参数和当前参数同时进行球场检测. 设这两组模型参数在同一帧图像中检测到的球场颜色像素比例分别是 R_i 和 R_c , 如果他们都大于阈值 T_d , 并且 $|R_i - R_c|$ 大于阈值 T_r , 这时把模型当前参数 Θ_c 恢复为初始参数 Θ_i , 其中 $T_d = 0.5$, $T_r = 0.1$. 前者保证一帧图像中含有足够多的球场像素才判断模型参数是否漂移, 后者则要求模型的初始参数和当前参数在检测结果上不应该差别很大.

3 球场检测在足球视频分析中的应用

球场检测算法在足球视频分析中有着重要作用, 本节利用足球球场检测的结果对足球比赛节目中的场景进行分类, 包括: 将足球视频序列分为场外场景、场内场景; 场内场景又分为特写和全局场景; 对于全局场景, 又进一步检测球门附近区域, 为下一步语义事件检测(如检测射门)提供线索. 图 3 描述了场景分类过程.

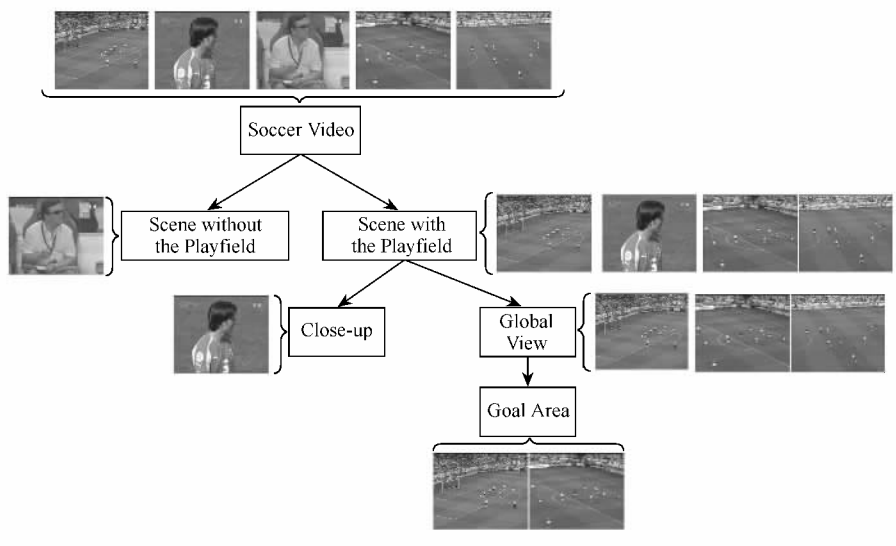


Fig. 3 The process of scene classification.

图 3 场景分类过程

3.1 球场内和球场外的判别

当一幅图像中球场区域的比例小于一定比例(本文取 0.05)时,应判别为场外区域。通常这类镜头用来描写观众席和教练席,同时也对应着非比赛片断。

3.2 球场内场景的分类

对于由上一步判别为球场内的场景,又进一步被分为含有球场区域的特写场景和全局场景。后者通常是视频流主要组成部分,数量巨大。由于足球

视频的精彩片断多发生在球门区域附近,所以从全局场景中进一步找到球门区域可以压缩掉大量的比赛过程,为生成摘要提供帮助。

本文提出用直线拟合球场区域在图像中的上边缘和下边缘,并根据拟合线内非球场区域比例和拟合线的斜率分类场内特写、球门区域和非球门区域。其过程如下:

1) 球场边界检测和拟合

图4(a)和图4(b)分别描述了中远场景和场内

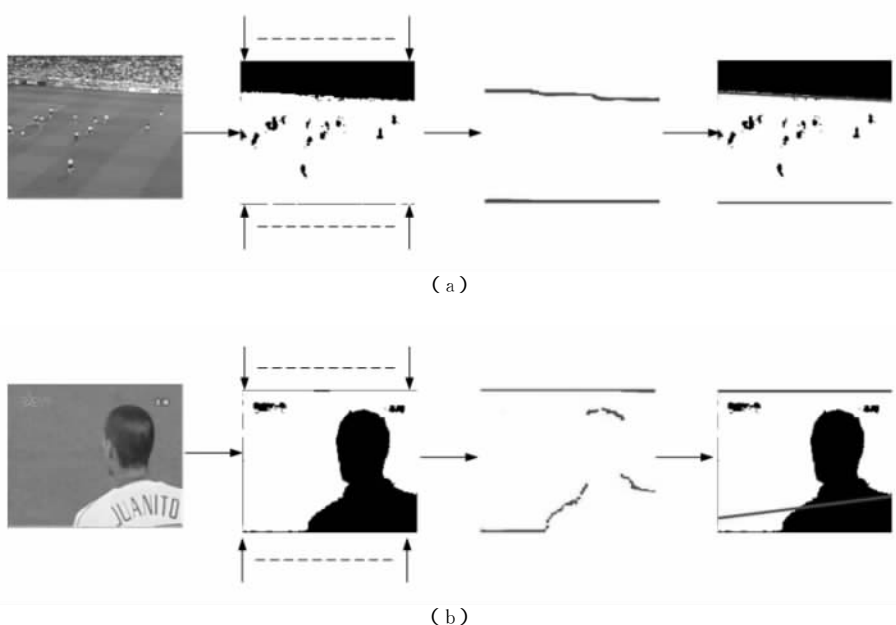


Fig. 4 The extracted brims of the playfield from images. (a) The case of global view; and (b) The case of a close-up for a player.

图 4 图像中球场边缘提取。(a) 中远场景 (b) 背景为球场的特写

特写镜头中的球场边缘拟合过程. 其中第 1 列为原始图像, 第 2 列为球场检测后的二值图像, 第 3 列为检测到的边缘点. 其过程是, 首先用垂直图像的射线从上向下、从左到右扫描二值图像, 将每一列中碰到的第 1 个球场像素点作为一个上边缘点 e_i^{up} , 则可以得到上边缘点集 $\{e_1^{up}, \dots, e_m^{up}, m \leq \text{图像宽度}\}$. 同理, 如果射线从下向上扫描图像, 则可以得到下边缘点集 $\{e_1^{bottom}, \dots, e_n^{bottom}, n \leq \text{图像宽度}\}$. 此时, 分别用两条直线拟合这两个集合, 得到两条直线 $y_i = K_i x + B_i, i = 1, 2$. 第 4 列为得到的拟合直线.

2) 场内特写和全局场景分类

从图 4(a) 和图 4(b) 的最后一列图中可以看到, 两条拟合直线和图像的边构成一个封闭区域, 设 MA 为封闭区域内最大连通的非球场区域面积, A 为这个封闭区域的面积. 如果

$$\frac{MA}{A} > T_a,$$

则认为场内特写, 否则为全局场景, 因为全局场景中, 球员和球在图像中所占区域很小. 本文中 T_a 设置为 0.25.

3) 球门区域检测

对于全局场景, 由于球门区域在主相机拍摄的图像中呈现出图 5 所示的模式, 则依据其上边缘的拟合直线斜率 K_1 判断左右球门区域, 判断准则如下:

$$S_q = \begin{cases} \text{左侧球门区域, } & K \geq T_{left}, \\ \text{右侧球门区域, } & K \leq T_{right}, \end{cases}$$

其中 T_{left} 和 T_{right} 是判断左右两个球门区域的两个阈值, 在两者之间被认为是中间区域, 本文中采用的



Fig. 5 Four general patterns of goal area appear in soccer video (From left to right: left-up, left-bottom, right-bottom and right-up positions in the court).

图 5 球门附近区域在主相机中呈现出的模式(前两个为左侧, 后两个为右侧)

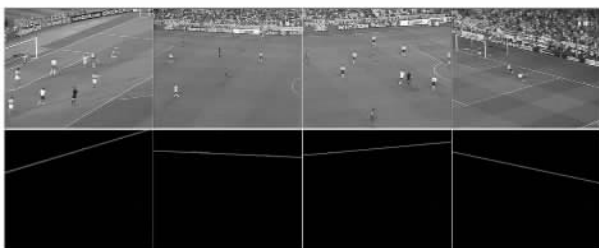


Fig. 6 Classification results of playfield region.

图 6 球场区域分类结果

值为 ± 0.1 . 图 5 给出球门区域呈现的几种模式. 图 6 给出了一些在实验中得到的结果.

4 实验及结果分析

为了检验本文提出的算法性能, 我们在实际数据上进行了大量测试. 这些数据录自于中央电视台体育频道和北京电视台体育频道, 包括足球、羽毛球、篮球、网球等比赛, 总长度超过 20h.

4.1 自适应高斯混合模型的检测结果

为了能够量化评价算法, 作者手工标注了部分视频片断, 包括 3 段足球以及 1 段篮球、1 段羽毛球、1 段网球, 每段视频均匀采样了 20 帧图像. 表 1 给出了高斯混合模型、自适应高斯混合模型以及直方图方法对球场检测的准确率. 从中可以看到, 利用自适应高斯混合模型进行球场检测取得了更好的结果. 其中, 精度定义为

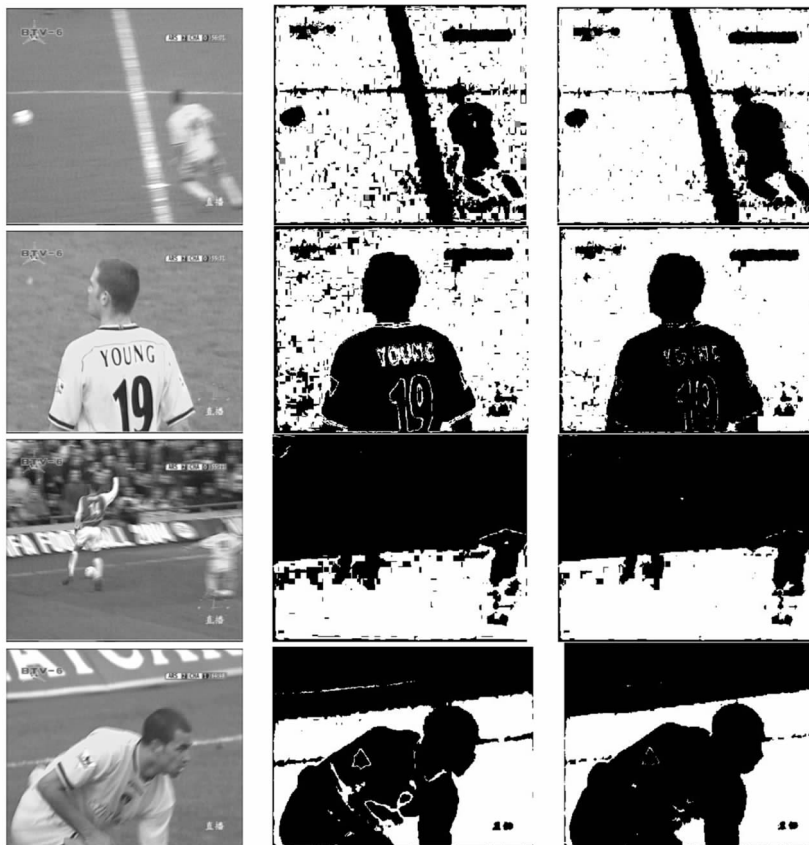
$$\text{精度} = \frac{\# \text{正确标记的球场像素和非球场像素的个数}}{\# \text{全部像素个数}}$$

Table 1 The Comparison of Detection Accuracy of Different Methods

| 表 1 检测结果的精度比较 | | | | % |
|----------------|-----------|-------|--------------|---|
| Video sequence | Histogram | GMM | Adaptive GMM | |
| Soccer-1 | 80.81 | 84.56 | 90.62 | |
| Soccer-2 | 90.76 | 95.48 | 98.96 | |
| Soccer-3 | 93.33 | 96.42 | 99.03 | |
| Badminton | 94.83 | 94.67 | 98.00 | |
| Tennis | 93.45 | 95.27 | 96.85 | |
| Basketball | 90.63 | 94.65 | 98.17 | |

图 7 给出了在一图像序列中的球场检测结果 (没有进行滤波处理). 其中第 1 列是图像序列, 第 2 列是用高斯混合模型检测的结果, 第 3 列是用自适应高斯混合模型检测的结果; 时间轴是从上到下. 从图中可以看到, 基于自适应高斯混合模型的球场检测取得了更好的效果. 这是因为在系统运行过程中其不断地修正模型参数, 使模型可以更好地描述整个视频流中球场颜色分布.

图 8 给出了基于自适应高斯混合模型的球场检测算法在不同体育视频中的检测效果图, 包括篮球、羽毛球、网球以及球场非常差的视频. 其中第 3 行给出了球场上有阴影以及比赛场地比较差的情况下的检测结果.



First column :original images. Results detected by GMM and adaptive GMM are in the second and the third column (without post processing)

Fig. 7 Results comparison.

图 7 基于高斯混合模型和基于自适应高斯混合模型的球场检测方法的结果对比



Fig. 8 The detection results by adaptive GMM in different sports videos.

图 8 自适应高斯混合模型在不同体育视频中的球场检测结果(未经过滤波处理)

4.2 足球视频场景分类实验

在基于自适应高斯混合模型的球场检测结果上,对第3节中提出的足球视频场景分类方法进行了性能测试.测试数据来自于2004年欧洲杯上的6场比赛,共有近810min的视频节目.表2给出了在6场比赛上的分类结果的查全率和准确率,它们分别定义为

$$\text{查全率} = \frac{\# \text{ 正确分类的场景个数}}{\# \text{ 此类场景个数}},$$

$$\text{查准率} = \frac{\# \text{ 正确分类的场景个数}}{\# \text{ 被分为此类的场景个数}}.$$

Table 2 Scene Classification for Soccer Video

(r is the recall and p is precision)

表2 足球视频场景分类结果(r 表示查全率, p 表示查准率)
%

| Match | Scene without playfield(r/p) | Close up (r/p) | Global View (r/p) | Goal area (r/p) |
|-------------------|-------------------------------------|-----------------------|--------------------------|------------------------|
| England :France | 100/100 | 94.7/90.2 | 100/89.8 | 95.7/91.5 |
| Portugal :Spain | 100/100 | 95.3/91.1 | 98.6/91.7 | 93.3/89.9 |
| Greece :Russia | 100/100 | 96.6/89.7 | 100/92.1 | 94.6/89.3 |
| Italy :Sweden | 100/100 | 94.1/89.2 | 97.4/90.3 | 93.7/90.6 |
| Germany :Holland | 100/100 | 95.8/92.5 | 98.2/85.2 | 92.9/87.6 |
| Denmark :Bulgaria | 100/100 | 92.3/93.4 | 96.9/88.9 | 94.1/91.5 |



Fig. 9 The unsolved problem : shadow.

图9 尚未解决的严重阴影球场

参 考 文 献

- 1 M. Luo, Y. F. Ma, H. J. Zhang. Pyramid wise structuring for soccer highlight extraction [C]. IEEE Pacific-Rim Conf. Multimedia, Singapore, 2003
- 2 A. Ekin, A. M. Tekalp. Robust dominant color detection and color-based application for sports video [C]. IEEE Int'l Conf. Image Processing, Barcelona, Spain, 2003
- 3 S. Q. Jiang, Q. X. Ye, W. Gao. A new method to segment playfield and its applications in match analysis in sports video [C]. ACM Multimedia, New York, NY, 2004
- 4 L. Wang, B. Y. Zeng, S. Lin, et al. Automatic extraction of

从实验结果上可以看到,本文第3节中提出的足球视频场景分类方法是有效的,而且它可以非常容易地扩展到曲棍球等其他体育项目的视频分析.

5 结论及今后工作

本文首先提出了一个基于自适应高斯混合模型的球场检测算法.它的最大特点是通过利用增量期望最大算法(IEM)在系统运行过程中,对高斯混合模型的参数不断调整.这种方法至少带来了如下优点:

1) 在线完善模型,使其适应视频流中球场颜色变化;

2) 学习过程中无需保留训练样本,节省了存储空间.

实验表明,本算法框架是可行的,而且相对于传统方法取得了更好的结果.

其次,本文提出的足球场景分类方法,有效地利用了球场区域在图像中的分布,方法简单有效.

在很多体育比赛中,由于太阳和灯光的照射,造成球场上留下球员的影子或者看台的影子,这给依赖于球场信息进行分析视频的方法带来不良影响(如图9所示).我们未来的工作就是要结合颜色模型和区域分析,解决球场检测中的阴影问题.

semantic colors in sports video [C]. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, Quebec, Canada, 2004

- 5 L. X. Xie, P. Xu, S. F. Chang, et al. Structure analysis of soccer video with domain knowledge and hidden markov models [J]. Pattern Recognition Letters, 2004, 25(7): 767~775
- 6 P. Xu, L. Xie, S. F. Chang, et al. Algorithms and systems for segmentation and structure analysis in soccer video [C]. IEEE Int'l Conf. Multimedia and Expo, Tokyo, Japan, 2001
- 7 A. Ekin, A. M. Tekalp. Shot type classification by dominant color for sports video segmentation and summarization [C]. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, Hong Kong, 2003
- 8 O. Utsumi, K. Miura, I. Ide, et al. An object detection method for describing soccer games from video [C]. IEEE Int'l Conf.

Multimedia and Expo, Lausanne, Switzerland, 2002

- 9 X. Yu, Q. Tian, K. W. Wan. A novel ball detection framework for real soccer video [C]. IEEE Int'l Conf. Multimedia and Expo, Baltimore, USA, 2003
- 10 D. Yow, B. L. Yeo, M. Yeung, *et al.* Analysis and presentation of soccer highlights from digital video [C]. Asian Conf. Computer Vision, Singapore, 1995
- 11 J. Assfalg, M. Bertini, C. Colombo, *et al.* Semantic annotation of soccer videos: Automatic highlights identification [J]. Computer Vision and Image Understanding, 2003, 92(2/3): 285~305
- 12 R. M. Neal, G. E. Hinton. A new view of the EM algorithm that justifies incremental and other variants [G]. In: Learning in Graphical Models. Cambridge, MA: Kluwer Academic Publishers, 1998. 355~368
- 13 Zhang Yujin. Image Engineering: Image Processing and Analysis [M]. Beijing: Tsinghua University Press, 1999. 179~180 (in Chinese)
(章毓晋. 图像工程(上册) 图像处理和分析 [M]. 北京: 清华大学出版社, 1999. 179~180)



Liu Yang, born in 1976. Received the B.A.'s and M.A.'s degrees in computer science and technology from Harbin Institute of Technology, in 1999 and 2001 respectively, where he is currently working towards

Ph.D. degree. His research interests include image processing, computer vision and machine learning.

刘扬, 1976年生, 博士研究生, 主要研究方向为图像处理、计算机视觉、机器学习。



Huang Qingming, born in 1965. Ph.D., professor and Ph.D. supervisor. His main research interests are in the areas of multimedia technology, video coding and compression, image processing and computer vision.

黄庆明, 1965年生, 博士, 教授, 博士生导师, 主要研究方向为多媒体技术、视频编码和压缩、图像处理和计算机视觉等。



Gao Wen, born in 1956. Received his Ph.D. degrees from the Harbin Institute of Technology in 1988 and the University of Tokyo in 1991, respectively. Now he is professor and Ph.D. supervisor in the

Institute of Computing Technology, Chinese Academy of Sciences, senior member of CCF. His main research interests include multimedia data compression, image processing, computer vision, multimodal interface, artificial intelligence and virtual reality.

高文, 1956年生, 博士, 教授, 博士生导师, 中国计算机学会高级会员, 主要研究方向为多媒体数据压缩、图像处理、计算机视觉、多模式接口、人工智能和虚拟现实。



Ye Qixiang, born in 1978. Ph.D. candidate of the Institute of Computing Technology, the Chinese Academy of Sciences. His main research interests are in the areas of image processing, pattern recognition and

multimedia application.

叶齐祥, 1978年生, 博士研究生, 主要研究方向为图像处理、模式识别和多媒体应用技术。

Research Background

Over the past decade, we have seen the explosive growth in the amount of available multimedia information in our daily lives. This trend necessitates the development of content-based video indexing and retrieval techniques. Since humans tend to use high level semantic concepts when querying and browsing multimedia databases, it is critical to develop techniques for semantic video analysis. In recent years, extensive research efforts have been devoted to sports video content analysis and applications due to their wide viewership and high commercial potentials. Technologies and prototypes have been developed to automatically or semi-automatically analyze sports video content, extract semantic events or highlights, intelligently adapt, enhance and personalize the content to meet users' preferences and network/device capabilities. Many techniques have been developed and used in broadcasting video analysis enhancement such as object detection and tracking, 3D virtual sports events reconstruction, scene analysis, audio classification, event detection and summarization, etc.