

# 基于视觉注意模型和 HMM 的足球视频语义分析

于 舟 张 瑞 杨小康

(上海交通大学电子工程系图像通信与信息处理研究所, 上海 200240)

(上海交通大学上海市数字媒体处理与传输重点实验室, 上海 200240)

**摘 要** HMM 模型具有良好的适应性, 可以自动学习, 对预测随机时序数据性能良好。场景是足球视频的基本特征, 场景的转换体现了足球视频的摄制、编辑模式, 表现了足球视频的语义。提出了一种基于场景分析和 HMM 的视频语义分析框架, 用于识别足球视频中的一些语义事件。为了克服以往基于主颜色和其他底层特征的视频场景分析中存在的较大误差, 又提出基于视觉注意模型对足球视频中的场景进行分析。实验结果表明, 基于场景分析和 HMM 的事件识别方法对足球视频中的任意球事件有良好的识别效果

**关键词** 语义分析 事件识别 图像要旨 HMM 场景分析

**中图法分类号**: TP301.6 **文献标识码**: A **文章编号**: 1006-8961(2008)10-2031-04

## Semantic Analysis of Soccer Video Using Model of Visual Attention and HMM

YU Zhou, ZHANG Rui, YANG Xiao-kang

(Institute of Image Communication and Information Processing, Department of Electronic Engineering,  
Shanghai Jiaotong University, Shanghai 200240)

(Shanghai Key Laboratory of Digital Media Processing and Transmissions, Shanghai Jiaotong University, Shanghai 200240)

**Abstract** A hidden Markov model (HMM) has good adaptability, can automatically learn and adapts well when used in predicting temporal data. Changing of scenes in soccer video is a fundamental trait and manifests semantics of soccer video. In this paper, a new method based on scene classification and HMM is given to analyze semantics of video, detecting semantic events in soccer video. In order to solve the low accuracy and low robustness of scene classification based on traditional method, model of visual attention and gist of scene is used in classification of scenes of soccer video. The experimental results show higher efficiency and higher accuracy of our method in detecting free kick in soccer video.

**Keywords** semantic analysis, event detection, GIST, HMM(hidden Markov model), scene classification

## 1 引 言

足球视频中存在一些特定的语义事件, 如任意球、角球、点球和越位等, 可以通过语义分析来对这些事件进行检测。在事件检测中, 由于隐马尔可夫模型(hidden markov model, HMM)具有适应性好、可自动学习、对预测随机时序数据性能良好等特点而被广泛使用。哥伦比亚大学的 Xie 等人基于隐马尔

科夫模型(HMM)分析足球的语义结构<sup>[1]</sup>, Ekin 等人进一步对足球视频进行了分析<sup>[2]</sup>。实验都表明, HMM 有效且稳定。

常见的 HMM 语义分析和事件识别算法提取一些底层特征(如主颜色等)作为 HMM 的观察值, 然而这些特征不能充分反映足球视频的特点, 使得算法准确性较差。同时, 这些算法均基于一些领域知识<sup>[3]</sup>, 例如足球领域知识, 适应范围过于狭窄, 面对比赛场地、光照的变化, 鲁棒性不高等。

**基金项目**: 国家自然科学基金委项目(60502034, 60625103); 国家 863 计划项目(1006AA01Z124)

**收稿日期**: 2008-07-12; **改回日期**: 2008-07-31

**第一作者简介**: 于 舟(1984 ~ ), 男。上海交通大学电子工程系图像通信与信息处理研究所硕士研究生。主要研究领域为视频语义分析和事件检测。E-mail: zhoyuyz@sjtu.edu.cn

场景是足球视频拍摄结构的基本单位,场景的转换充分表现了足球视频的语义。本文提出基于场景变换分析足球视频,并将视觉注意模型运用到足球视频的场景分析中<sup>[4,5]</sup>。实现了基于场景分析和 HMM 的足球视频语义分析,有效提高了足球视频语义事件识别的准确率。

2 足球视频的场景分析

2.1 基于视觉注意模型的足球视频场景分析

以往的足球场景识别基于主颜色分析,准确率不高,鲁棒性差。本文给出了一种基于视觉注意模型和图像要旨的足球场景识别方法,用全局特征向量——图像要旨作为场景识别的标识。

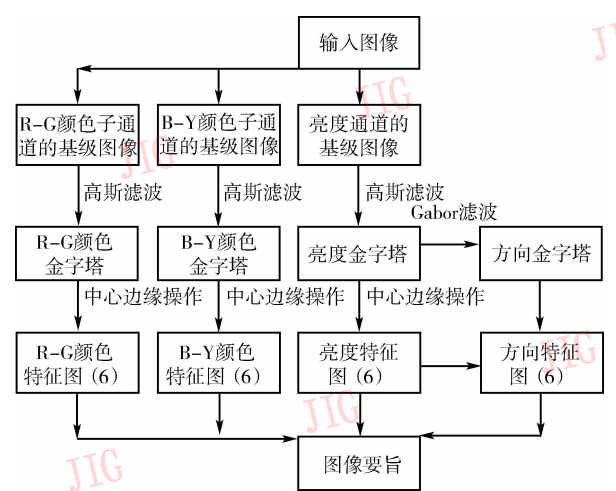


图 1 图像要旨模型框图  
Fig. 1 Structure of gist

每一幅输入的足球视频帧图像都被分解到亮度、颜色和方向 3 个通道上,然后通过中心-边缘操作或者 Gabor 滤波生成一系列特征图  $M_i$ 。特征图个数为  $1 \times 6 + 2 \times 6 + 4 \times 4 = 34$ 。

对每一个特征图,本文将其划分为  $4 \times 4$  块,计算每一块的平均值,作为图像要旨向量。

$$G_i^{k,l} = \frac{1}{16WH} \sum_{u=\frac{kW}{4}}^{\frac{(k+1)W}{4}-1} \sum_{v=\frac{lH}{4}}^{\frac{(l+1)H}{4}-1} M_i \tag{1}$$

式中,  $i$  表示特征图的序号,  $k$  和  $l$  表示图像子块的坐标,  $W$  和  $H$  分别表示整幅图像的宽和高。

将 34 个特征图的图像要旨向量拼接起来,形成全图图像要旨向量,维数为  $34 \times 16 = 544$  维。

本文基于 SVM 对特征向量进行分类,将足球视频场景分为全局、局部、特写和场外 4 种场景。测试

30 小时的足球视频,效果良好,如表 1 所示。

表 1 场景分类的准确率  
Tab. 1 The accuracy of scene classification

| 场景  | 特写        | 全局        | 局部        | 场外        |
|-----|-----------|-----------|-----------|-----------|
| 准确率 | 0. 933 42 | 0. 955 98 | 0. 929 78 | 0. 828 16 |

图像要旨的提取采用了图像金字塔的多分辨率处理方法,因此,对于缩放、旋转、平移具有良好的不变性。这样足球视频中拍摄的角度、位置、远近都无法对场景的分类造成大的负面影响。

图像要旨能抵抗光照带来的影响。光照对足球场景的影响很大,足球场景的不同部分也存在互相阴影的遮挡。由于预先将图像分割到亮度和颜色通道,采用了颜色对处理,加之中心-边缘操作,图像要旨具有一定程度的关照不变性,大大提高了足球场景分类对于光照的鲁棒性。

2.2 其他特征的提取和全局场景的细分

2.2.1 足球场地标志线

足球场地图标志线可以将全局场景细分为前场或后场、中场和无法判断。如果有平行而且斜率较小的标志线就是前场或后场,如果有斜率接近 90 度的单根标志线就是中场,其余为无法判断。检测标志线的算法为霍夫(Hough)变换。

2.2.2 场地边缘检测

场地边缘的不同走向可以区分全局场景是在对方半场还是在本方半场。由于足球场地图边缘明暗变化较为复杂,直接进行边缘检测效果很差。本文提出在检测出边缘后用 Hough 变换判断场地边缘的大致走向,进而判断场地是在对方半场还是在本方半场。

2.2.3 场景的全局运动

场景的全局运动可以区分场景是进攻还是防守。本文利用块匹配算法来检测场景全局运动。

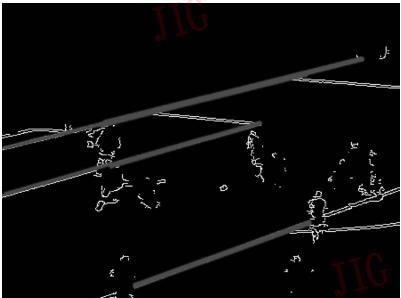


图 2 场地标志线和场地边缘检测  
Fig. 2 Detection of line and edge of field

### 3 足球视频语义分析和事件检测框架

#### 3.1 足球视频的结构分析

根据足球视频的特点,本文将足球视频的场景分为全局、局部、特写和场外 4 类,并对全局场景进行了细分。本文使用颜色直方图和角点进行镜头分割。

#### 3.2 任意球的拍摄模式和结构分析

足球视频中任意球事件有固定的摄制、编辑模式。首先,在犯规后,几乎总有一个全局场景表现球员落位,标志任意球事件开始。如图 3 所示。

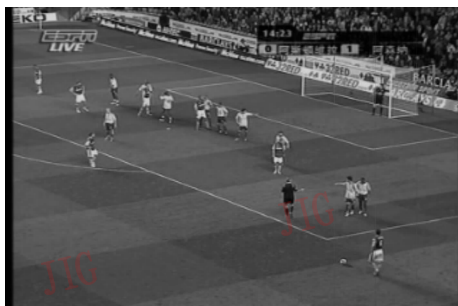


图 3 任意球场景一

Fig. 3 First step of free kick

其次,会有 1 到 3 个场景表示球员的抢位和任意球的准备,如图 4 所示。



图 4 任意球场景二

Fig. 4 Second step of free kick

最后,会有一个全景场景(发球),如图 5 所示。



图 5 任意球场景三

Fig. 5 End of free kick

本文设定训练和检测的任意球视频都从一个前场全局场景开始,由一个前场全局场景结束,中间由 1 到 3 个过渡场景组成,它们的状态是任意的。实际的任意球视频中符合以上规律的占的 85% 以上。

#### 3.3 任意球事件的 HMM 模型构建

##### 3.3.1 HMM 概述

HMM 模型的基本原理如下<sup>[6]</sup>:

对于一个随机事件,有一个观察值序列:  $O_1, O_2, \dots, O_T$ , 隐含一个状态序列:  $X_1, X_2, \dots, X_T$ 。

假设 1: 马尔可夫假设 (构成一阶马尔可夫链)

$$p(X_i | X_{i-1}, \dots, X_1) = p(X_i | X_{i-1}) \quad (2)$$

假设 2: 不动性假设 (状态与具体时间无关)

$$p(X_{i+1} | X_i) = p(X_{j+1} | X_j) \quad (3)$$

假设 3: 输出独立性假设 (仅与当前状态有关)

$$p(O_1, \dots, O_T | X_1, \dots, X_T) = \pi p(O_1, \dots, O_T) \quad (4)$$

上式中,  $i, j$  表示状态值和观察值的序号,  $\pi$  表示初始概率分布。

##### 3.3.2 Markov 链的构建

为了对任意球事件进行语义分析,必须确定任意球事件的状态和 Markov 链的形状。

本文设定 HMM 状态如下:

(1) 全局场景 (对方前场的全局场景为任意球的第 1 个和最后一个场景)

(2) 局部场景

(3) 特写场景

(4) 非球场区域场景

根据视频训练集中任意球事件的场景转换序列,可以确定 Markov 链的形状。一种任意球场景转换序列可以抽象为一种状态转移序列。例如,状态(1)-状态(3)-状态(2)-状态(1)。

总结训练集中任意球视频的状态转换方式,任意球事件的状态转移规律为从状态(1)开始,经过(1)~(3)任意转换的状态,由状态(1)结束。

##### 3.3.3 HMM 观察值

状态的描述向量,即 HMM 模型的观察值由如下部分构成:

(1) 全局场景,局部场景,特写场景,场外场景

(2) 前场场景,中场场景,后场场景

(3) 对方半场,本方半场

(4) 进攻、防守 (摄像机向对方半场移动,向本方半场移动)

其中,(2)、(3)、(4)只与全局场景有关。

3.3.4 HMM 的训练

在完成了 HMM 模型的设计和场景描述向量的提取之后,就可以对 HMM 进行训练,确定 HMM 的参数  $\lambda = (\pi, A, B)$ 。

本文的训练集均为截取的任意球视频。包括 12 场英超 Arsenal 对其他球队的比赛,比赛总长度超过 20 小时。

状态转移概率矩阵  $A$  表示从某一时刻某种状态转换到下一时刻下某种状态的概率,这里通过 Baum-Welch 算法求得状态转移矩阵。

观察值转移概率矩阵  $B$  的计算方法如下:

$$b_{ij} = \frac{n_{ij}}{n_i} \quad i, j \in N \tag{5}$$

式中,  $n_{ij}$  表示从状态  $i$  观察到观察值  $j$  的场景数,  $n_i$  表示从状态  $i$  得到各种观察值的总数。

本文假定所有任意球视频都从状态 1 开始。初始概率分布  $\pi$  固定。

3.3.5 任意球事件的识别

在 HMM 模型参数为  $\pi, A, B$  的基础上,可以对任意球视频进行识别,过程如下:

- (1)对足球视频进行镜头分割和关键帧提取。
- (2)对关键帧进行场景分析。
- (3)基于 HMM 模型和观察值序列,通过前向-后向算法解码,得到状态序列的概率分布。
- (4)通过 HMM 判决函数,判决是否为任意球。

3.3.6 任意球事件的判决函数

HMM 的前向-后向算法可以在给定模型参数和观测序列的情况下,快速求出状态序列发生的概率。在 HMM 判决函数中,预先设定了一些序列作为任意球序列(这些序列从训练集中取得)。分别计算这些序列发生的概率,把其中最大的概率与阈值比较,如果大于阈值则认为此序列为任意球事件,小于阈值则认为不属于任意球事件。

4 实验结果和分析

基于视觉注意模型进行场景分类,特写、全局、局部 3 类场景的分类准确率在 90% 以上,场外场景在 80% 以上,如表 1 所示。

基于场地标志线、场地边缘、全局场景运动的全局场景细分准确率均在 80% 以上。

根据本文的算法,对任意球事件和非任意球事件进行了实验,测试集中包括 8 场英超 Arsenal 对其他球队的比赛,长度超过 15 小时。采用查准率和查全率来对结果进行评价,其中,查准率表示检测得到正确的任意球事件占检测得到的所有事件的比例,查全率表示检测得到正确的任意球事件占有任意球事件的比例。实验结果如表 2 所示。

表 2 任意球识别准确率  
Tab. 2 The accuracy of detection of free kick

| 单位: % |      |     |
|-------|------|-----|
|       | 查准率  | 查全率 |
| 任意球   | 78.2 | 77  |

5 结 论

本文提出了一种基于场景分析和 HMM 的视频语义分析框架。利用视觉注意模型和图像要旨对足球视频场景进行分类,并通过其他特征进一步细分场景。基于 HMM 对足球视频的场景转换建模,识别足球视频中一些典型的语义事件,例如任意球。实验结果表明,基于 HMM 和场景分析的事件识别方法对任意球事件有良好的识别效果。

参考文献 (References)

1 Xie L, Chang S F, Divakaran A, et al. Structure analysis of soccer video with hidden markov models [ A ], In: Proceedings of International Conference on Acoustic, Speech and Signal Processing [ C ], Orlando, FL, USA, 2002:13 ~ 17.

2 Ekin A, Tekalp A M, Mehrotr R. Automatic soccer video analysis and summarization [ J ]. IEEE Transactions on Image processing, 2003, 12(7):796 ~ 807.

3 Zhong Di, Shi-fu Chang. Structure Analysis of sports video using domain models [ A ]. In: Proceedings of IEEE International Conference on Multimedia and Expo [ C ], Tokyo, Japan, 2001: 920 ~ 923.

4 Turk M, Pentland A. Eigenfaces for recognition [ J ]. Journal of Cognitive Neuro-science, 1991, 3(1):71 ~ 86.

5 Penev P S, Atick J J. Local feature analysis: A general statistics theory for object representation[J]. Network: Computation in Neural Systems, 1996, 7(3): 477 ~ 500.

6 Rabiner Lawrence R. A tutorial on hidden Markov models and selected applications in speech recognition[J]. Proceedings of IEEE, 1989, 77(2): 257 ~ 286.