

网球视频分析的运动向量场变换算法

王 鹏, 蔡 锐, 李 彬, 杨士强

(清华大学计算机科学与技术系, 北京 100084)

摘 要: 网球比赛中摄像机与球场平面成一倾角, 利用运动估计难以获得球员的真实跑动信息, 且容易受到大量随机噪声干扰. 本文基于针孔摄像机模型, 提出一种运动向量场变换算法, 提高基于运动向量场特征分析网球视频的性能. 该算法利用前景物体掩蔽和全局运动补偿技术, 去除随机噪声干扰. 实验结果验证了本文算法的有效性和稳定性.

关键词: 网球视频分析; 运动向量场变换算法; 针孔摄像机模型

中图分类号: TP391.43 **文献标识码:** A **文章编号:** 0372-2112(2005)05-0935-04

Tennis Video Analysis Using Transformed Motion Vector Field

WANG Peng, CAI Rui, LI Bin, YANG Shi qiang

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: The motion estimated from tennis video is unable to analyze player's true movement due to the diagonally located camera and the high noises. In this paper, a transform algorithm of Motion Vector Field (MVF) is proposed to revise the deformed motion vector based on the pinhole camera model. In this algorithm, the techniques of foreground object mask and global motion compensation are incorporated to reduce noises. Previous methods based on MVF analysis are further utilized to recognize basic actions of either tennis player. And, the experiments on live broadcasted tennis videos show very encouraging results.

Key words: tennis video analysis; transform algorithm of motion vector field; pinhole camera modeling

1 引言

运动特征是视频媒体中最为重要也是目前研究最多的图像特征之一, 从视频压缩流中提取的运动向量反应两帧图像中对应宏块的位移. 运动向量组成的稠密场构成了帧图像的运动向量场. 目前体育视频分析广泛使用运动向量场的特征. 例如, Duan 等^[1]提出基于运动向量场的中间层表征框架, 并在多种体育视频中应用. Ma 在文献[2]提出运动能量谱算法并用于视频检索, 而文献[3]面向语义事件检测提出运动能量重分布的度量. 前人工作表明基于运动向量场的体育视频分析具有速度快和性能鲁棒的优点.

然而, 在网球视频分析中, 基于运动向量场的分析工作很少. 传统的网球视频分析集中在检测和跟踪球员和网球的运动轨迹^[4,5], 以及识别挥拍手势和动作^[6]. 基于计算机视觉的方法可以对内容进行精细描述, 但其主要缺点是计算复杂度高、应用不灵活、局限性较大. 通过深入研究, 本文认为制约运动向量场在网球视频分析的两个主要原因是: (1) 由于摄像机拍摄与球场平面不垂直, 从视频帧序列估计得的运动向量难以反映球员的真实跑动信息. (2) 因为网球场颜色相似, 在估计宏块运动向量时容易产生大量随机噪声, 球员的运动向量很容易被随机噪声干扰. 为了在网球视频中应用基于运动向量场的分析方法, 必须要: 根据真实运动修正运动向量场和减少随机噪声的干扰. 本文提出一种运动向量场的变换算法,

该算法利用针孔摄像机模型对运动向量进行修正. 为了减少随机噪声的干扰以及摄像机在拍摄时轻微移动带来的扰动, 本文使用前景物体掩蔽和全局运动补偿技术作为预处理步骤. 为了验证变换算法的性能, 使用文献[3]的算法处理变换后的运动向量场检测网球运动员的基本技术动作. 通过对比实验, 充分验证本文变换算法的有效性.

2 运动向量变换

网球比赛中的摄像机通常放置在球场纵向对称轴上方, 拍摄角度与地面倾斜, 因此真实的矩形球场经拍摄后变形为等腰梯形球场, 如图1所示. 由视频序列估计的运动向量随之变形, 很难反映真实的球员运动情况. 图1左图给出网球视频的梯形球场示意图, 右图给出真实矩形球场示意图. 若视频序列的当前帧中一点 p_1 在下一帧对应点为 q_1 , 则运动向量可以

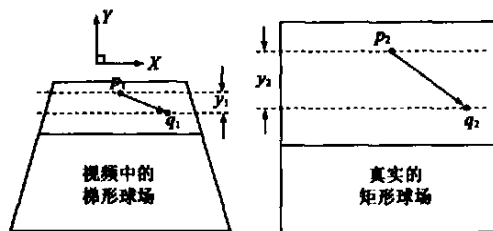


图1 运动向量的扭曲示意图

被表示为 p_1 到 q_1 的位移, p_2 和 q_2 分别是它们在真实球场中的对应点. 若从视频帧中看到球员从 p_1 运动到 q_1 , 则该球员在真实球场的运动是从 p_2 到 q_2 . 真实运动的大小和方向分别被缩小和扭曲, 通常都存在 $y_1 < y_2$. 这种变形使得分析球员的真实运动变得困难, 例如, 我们很难从运动向量在纵轴方向的投影直接判断球员是否上网.

对于视频梯形球场中的点 p_1 和 q_1 来说, 若能准确定位它们在真实矩形球场中的对应点 p_2 和 q_2 , 则可实现运动向量的修正. 因此, 问题转化为给定视频图像中任一点, 计算它在真实球场的对应点. 本文使用针孔摄像机模型, 如图 2 左上部分所示. 对于针孔摄像机, 有如下关系式, 其中, L 和 l 分别表示物体的长度和像的长度, u 和 f 分别表示物距和摄像机的焦距.

$$L/l = u/f \quad (1)$$

假定摄像机到球场下底线的水平距离为 d , 摄像机距离地面高度为 h , 由式(1)得到:

$$\begin{aligned} W/w_1 &= \sqrt{d^2 + h^2}/f \\ W/w_2 &= \sqrt{(H+d)^2 + h^2}/f \\ W/w_3 &= \sqrt{(2H+d)^2 + h^2}/f \end{aligned} \quad (2)$$

其中, W 和 H 分别表示真实球场的宽度和半长度^[9], w_1, w_2, w_3 分别表示视频中梯形球场的下底线, 中场线和上底线的长度, 示意图见图 2.

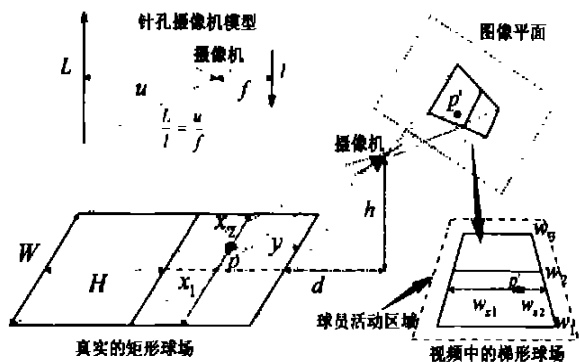


图 2 基于针孔摄像机模型的运动向量修正

图像平面内梯形球场中任意一点 p' , 经过 p' 与底线平行的直线被 p' 和两条边线分割为两部分, 其长度分别表示为 w_{x1} 和 w_{x2} . 假定 p 是 p' 点在真实矩形球场中的对应点, 而 p 由 x_1, x_2, y 唯一确定, 其中 x_1, x_2 分别表示 p 点到两条边线的距离, y 表示 p 点到下底线的距离. 根据针孔摄像机模型, w_{x1}, w_{x2} 和 x_1, x_2, y 之间有:

$$\begin{cases} W/(w_{x1} + w_{x2}) = \sqrt{(y+d)^2 + h^2}/f \\ w_{x1}/w_{x2} = x_1/x_2 \text{ 且 } x_1 + x_2 = W \end{cases} \quad (3)$$

由式(2)可求解出参数 d 和 h , 则利用式(3)可对图像平面中梯形球场内任意一点, 计算其在真实矩形球场中对应点. 通过对运动向量的两个端点分别进行点变换, 计算变换后的运动向量, 实现运动向量的修正. 在大多数的网球比赛视频中, 摄像机的位置基本满足本文假设. 利用现有的直线检测算法^[7], 检测视频帧序列中把梯形球场的边线、底线长度和角点坐标. 对每一个比赛镜头, 前几帧图像的球场检测结果取平均

值作为整个镜头的球场位置信息. 球员活动区域定义为比球场更大的一个梯形区域, 如图 2 中点划线梯形所示. 对球员活动区域内的全部运动向量, 将其变换到真实球场中运动向量.

3 前景物体掩蔽

由于网球场颜色一致, 基于宏块匹配的运动向量估计很容易带来随机噪声. 网球运动员和整个帧图像相比只占很小一部分, 因此分析球员运动时随机噪声的干扰不容忽视. 一种简单直接的方法是把球员区域(即前景物体)和背景区域分割开, 避免噪声影响. 此外, 摄像机有时会随球员的跑动而轻微平移, 使得运动估计结果包含全局运动分量. 尽管假设的摄像机模型仍然成立, 但补偿全局运动分量后才会获得更准确的信息.

前景物体分割和全局运动补偿技术在基于物体编码技术中广泛使用, 能用于解决上述问题. 本文采用文献[8]快速鲁棒的全局运动估计算法, 定位前景物体区域并补偿全局运动分量. 该算法使用六参数仿射模型表征两帧图像间的全局运动, 使用 Gauss-Newton 和 Levenberg-Marquardt 迭代算法估计参数. 在每个迭代步骤, 该算法使用残差宏块分离把属于前景物体的像素剔除, 修正全局运动估计的参数. 最终迭代结果得到前景物体掩蔽(Foreground Object Mask, FOM)和全局运动分量(Global Motion Component, GMC). 利用上述技术获得的前景物体掩蔽 FOM 和全局运动分量 GMC 和原始运动向量场维数相

同, 分别表示如下:

$$\begin{cases} \text{FOM} = (f_{i,j}) \\ \text{GMC} = (g_{i,j}) \end{cases} \quad (4)$$

其中 $f_{i,j}$ 的值为 0 或 1, 0 表示宏块 (i, j) 属于背景, 1 表示宏块 (i, j) 属于前景物体. $g_{i,j}$ 表示宏块 (i, j) 应补偿的全局运动分量. 图 3 给出一个实例, 左图是网球比赛视频的原始帧图像, 右图是使用算法^[8]获得带有前景物体掩蔽的图像. 实验中, 控制分离算法的阈值参数设为 10%.

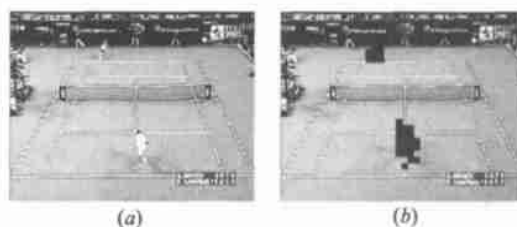


图 3 原始图像(a)和标有前景物体掩蔽的图像(b)

4 球员基本动作分类

图 4 给出运动向量场变换算法的流程图. 对于输入的网球比赛视频, 首先, 使用基于主颜色的算法^[4]把比赛镜头检测出来; 然后, 对每个比赛镜头提取原始运动向量场. 变换算法由两步骤组成, P 过程是预处理步骤, 用于计算前景物体掩蔽和全局运动分量. P 步骤中的 \odot 表示矩阵加法操作, \otimes 表示矩阵对应元素的乘法操作. T 过程是变换步骤, 利用球场检测结果, 对运动向量进行修正, 得到修正的运动向量场.

4.1 基于运动向量场提取时序运动特征

本文对球员的两类基本动作进行分类: 上网型和底线型. 上网动作定义为球员向球网移动并击球, 底线动作定义为球

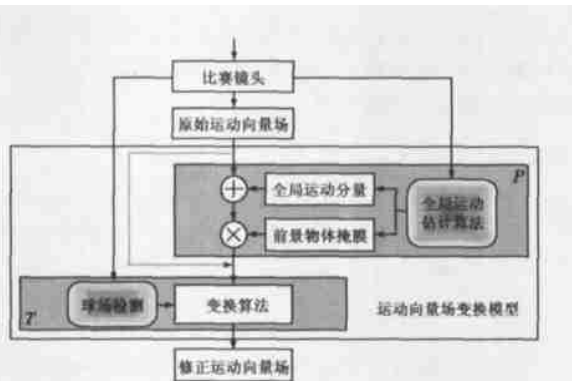


图 4 运动向量场变换模型的流程图

员在底线附近击球。利用文献[3]的基于运动向量场提取时序运动特征的方法, 实现上网/底线动作的分类, 简要介绍如下: 首先利用能量重分布函数计算能量场, 然后利用特定运动的权值模板与能量场卷积得到运动响应值。其中, 能量重分布函数量化运动向量的重分布, 是将块运动向量场转化为更容易计算的能量图谱。能量重分布仅依赖于当前帧中的块在参考帧中的匹配位置。如果用 $E_{x,y}^t$ 表示 t 帧中块 (x,y) 包含的能量, $\text{overlap}_{i,j,x,y}^t$ 表示 t 帧中块 (x,y) 与 $t-1$ 帧中块 (i,j) 在 t 帧中匹配位置的交叠面积, 块尺寸用 W_b 表示。则能量重分布函数定义如下:

$$E_{x,y}^t = \left[\sum_{i,j} \text{overlap}_{i,j,x,y}^t \times E_{x,y}^{t-1} \right] / W_b^2 (1 \leq i, j \leq W_b) \quad (5)$$

能量重分布函数提供一种帧间运动的表示形式, 权值模板本质是与块向量场尺寸一致的权值矩阵, 能从能量图谱序列中检出对应的运动。用不同的值、不同的排列顺序去填充权值模板中元素, 将能改变权值模板对不同运动的敏感性。如果用 $w_{x,y}$ 表示权值模板 (x,y) 处的元素, 使用如下公式获得给定能量图谱对该模板的响应值 E_{res}^t 。

$$E_{res}^t = \sum_{x,y} E_{x,y}^t \times w_{x,y} (1 \leq x, y \leq W_b) \quad (6)$$

实验中, 本文使用横向主运动和纵向主运动的权值模板, 对每个比赛镜头计算横向和纵向运动的响应值曲线, 并把响应曲线作为隐含马尔可夫模型的观测特征向量, 对两类动作分别建模。利用球场的中线位置把整个球场分为上下两部分, 分别对上下半场的球员动作分类。

为了验证本文的变换算法, 实验分别基于 (a) 原始运动向量场; (b) T 变换得到的运动向量场; (c) P 处理和 T 变换得到的运动向量场, 计算时域运动响应曲线。图 5 对为三种情况下上半场球员上网动作的相应曲线, 其中横坐标表示帧号, 纵坐标表示运动相应值。纵向运动响应值的正方向表示球员向近网方向运动, 横向响应值的正方向表示球员向右运动。从第 1 帧到第 80 帧, 球员从左底线向右侧近网方向运动, 从 81 帧到 134 帧, 球员转身从右侧跑回右底线。图 5(a) 中两条曲线抖动剧烈, 纵向运动曲线缺乏规律, 很难刻画上网动作。图 5(b) 中两条曲线响应范围均被拉大, 对上网动作和非上网动作的区分能力增强, 然而噪声带来的曲线抖动仍十分剧烈, 为分类带来困难。图 5(c) 中的曲线抖动明显减少, 纵向运动的曲线图基本能准确刻画上网动作, 更有利于分类。

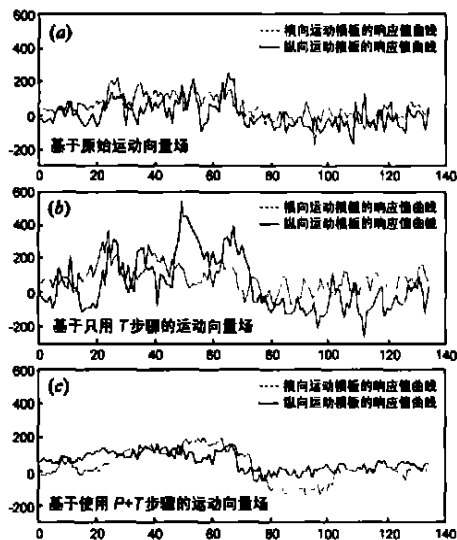


图 5 基于不同运动向量场生成的时域运动曲线图对比

4.2 基于 HMM 的上网/底线动作分类

对一个比赛镜头, 用两个隐含马尔可夫模型对上网和底线动作分别进行建模, 如图 6 所示。其中, 底线动作用一个状态的 HMM 描述, 其观测值用连续高斯混合模型 (GMM) 描述。由于底线动作大致由左跑、右跑、静止组成, 所以实验中高斯混合分量的个数为 3。上网动作用两状态的从左到右 HMM 描述, 其中一个状态描述近网跑动, 另一个状态描述返回底线。每个状态都用连续混合高斯模型描述, 其混合分量个数均为 3。两个 HMM 的观测向量是由横向和纵向运动响应值构成。

如图 6 所示, 两个 HMM 被连接成一个更高层的环形结构模型, 该模型有一个虚拟输入状态和一个虚拟输出状态, 分别表示在一个比赛镜头中底线动作和上网动作的转换。实验中, 从输入状态到两个子 HMM 的转移概率均为 0.5。通过手工标注获得上网/底线动作的训练数据, 对全部上网数据使用 EM 算法估计获得上网 HMM 模型参数, 反之对全部底线数据使用 EM 算法估计获得底线 HMM 模型参数。在识别阶段, 给定观测序列情况下 (即时序运动响应曲线), 使用 Viterbi 算法^[10]寻找最优全局状态路径, 实现两个动作的分类和分割。

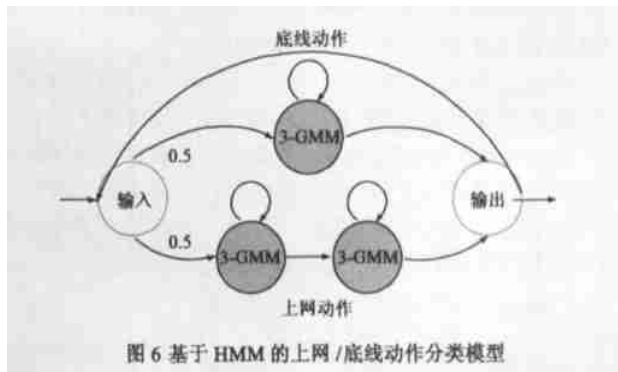


图 6 基于 HMM 的上网/底线动作分类模型

5 实验结果与分析

实验选取四个小时的网球比赛视频, 实验数据采集自 2002 年美国公开赛中 A. Agassi 和 P. Sampras 的比赛 (Video1)

和 2003 年温布尔登公开赛中 R. Federer 和 M. Philippoussis 的比赛 (Video2). 实验中对比赛镜头进行数据标注, 如果镜头包含球员的上网动作, 则把该镜头标注为上网镜头; 反之, 如果镜头不包含球员的上网动作, 把该镜头标注为底线镜头, 分别对上下半场的球员进行标注. 实验数据的具体信息见表 1.

表 1 测试视频数据的基本信息

视频	全部镜头	比赛镜头	上半场球员		下半场球员	
			上网	底线	上网	底线
Video1	881	316	58	258	230	86
Video2	624	271	100	171	114	157
Σ	1505	587	158	429	344	243

对某一半场的球员, 在识别过程, 如果比赛镜头中检测到上网片段, 则该镜头判定为上网比赛镜头, 反之判定为底线比赛镜头. 分别用原始运动向量场, 只用 T 变换得到的运动向量场, 使用 P 处理和 T 变换得到的运动向量场进行实验, 结果见表 2. 因为, 只用原始运动向量场时纵向运动响应曲线对上网

表 2 实验对比结果

运动 向量场	比赛 镜头	上半场球员		下半场球员	
		查准率(%)	查到率(%)	查准率(%)	查到率(%)
原始	上网	30.91	37.78	64.55	67.78
	底线	70.53	63.81	47.75	44.17
使用 T 变换	上网	61.11	73.33	85.06	82.22
	底线	87.50	80.00	74.60	78.33
使用 P 处理 和 T 变换	上网	80.00	66.67	90.85	82.78
	底线	86.67	92.86	77.21	87.50

和底线动作的区分能力很差, 见图 5(a). 加上噪音干扰使得曲线抖动过大, 很多底线动作误判为上网动作. 而且由于上半场的运动向量变形更严重, 对上半场球员的动作分类要比下半场的准确率低. 和图 5(a)相比, 图 5(b)中的横向运动和纵向运动响应值范围都被扩大, 使得利用运动特征区分上网动作和底线动作的能力大大提高. 因此, 基于 T 变换得到的运动向量场的分类性能明显提高, 尤其是上半场球员的上网镜头的查到率和查准率都提高了一倍多. 然而, 由噪声干扰引起的曲线抖动仍较为严重, 在寻找全局最优状态路径时, 造成较多的误检. 利用本文提出的 P 处理, 能有效减少噪声干扰, 图 5(c)中曲线更为平滑, 上网和底线动作的运动特征区别更加明显, 因此基于 P 处理和 T 变换得到的运动向量场的分类性能相对最高, 能够满足实际应用需要, 充分验证了本文变换算法的有效性. 在用宏块匹配算法估计原始运动向量场时, 引入的误差和错误运动向量会对实验性能造成影响. 然而, 从实验结果来看, 本文的变换算法对大多数的网球比赛镜头都能很好适用, 能有效修正运动向量场, 减少噪声干扰, 使得基于运动向量场的视频分析方法更有效地应用.

6 结语

本文面向网球视频分析提出一种基于针孔摄像机模型的运动向量场修正算法. 在该算法中, 根据球员的真实运动修正变形运动向量, 并使用前景物体掩蔽和全局运动补偿技术作为预处理步骤减少噪声干扰, 提高算法的稳定性和准确性. 实验对网球运动员的主要技术动作进行分类, 实验对比结果令

人满意. 后续工作可以: (1) 提高算法的鲁棒性, 例如快速准确的检测球场位置信息; (2) 利用音频特征和其他先验知识, 尝试更细致的球员动作分析.

参考文献:

- [1] L Y Duan, M Xu, et al. A mid-level representation framework for semantic sports video analysis[A]. Proc. of the 11th ACM International Conference on Multimedia[C]. Berkeley, CA, USA: 2003, Nov. 2-8, 33-44.
- [2] Y F Ma, H J Zhang. A new perceived motion based shot content representation[A]. Proc. of IEEE International Conference on Image Processing[C]. Thessaloniki, Greece: 2001, (13): 426-429.
- [3] G Xu, Y F Ma, et al. Motion based event recognition using HMM[A]. Proc. of the 16th International Conference on Pattern Recognition[C]. Quebec, Canada: 2002(2): 831-834.
- [4] G Sudhir, John C M Lee, Anil K Jain. Automatic classification of tennis video for high level content-based retrieval[A]. Proc. of 1998 International Workshop on Content-Based Access of Image and Video Databases[C]. Bombay, India: Jan. 1998, 03-03, 81-90.
- [5] G S Pingali, Y Jean, I Carlhom. Real time tracking for enhanced tennis broadcasts[A]. Proc. of 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C]. Santa Barbara, CA, USA: Jun: 1998, 260-265.
- [6] H Miyamori. Improving accuracy in behaviour identification for content-based retrieval by using audio and video information[A]. Proc. of the 16th International Conference on Pattern Recognition[C]. Quebec, Canada: 2002(2): 826-830.
- [7] H J Di, L Wang, G Y Xu. A three step technique of robust line detection with modified hough transform[A]. Proc. of SPIE Conference on Multispectral Image Processing and Pattern Recognition[C]. Beijing, China: 2003. 835-838.
- [8] Y W He, B Feng, et al. Fast global motion estimation for global motion compensation coding[A]. Proc. of IEEE International Symposium on Circuits and System[C]. Sydney, Australia: 2001. 233-236.
- [9] <http://www.hickoksports.com/glossary/gtennis.shtml>.
- [10] L R Rabiner, B H Juang. An introduction to hidden markov models[J]. IEEE Acoustics, Speech and Signal Processing Magazine, 1986, (3): 4-16.

作者简介:



王 鹏 女, 1979 年 1 月出生于北京, 2001 年毕业于清华大学计算机科学与技术系, 现为清华大学计算机系人机交互与媒体集成研究所博士研究生, 主要从事多媒体信息的基于内容描述、视频分析和索引等研究. E-mail: wangp01@mails.tsinghua.edu.cn.

蔡 锐 男, 1978 年 10 月出生于四川省绵阳市, 2001 年毕业于清华大学计算机科学与技术系, 现为清华大学计算机系人机交互与媒体集成研究所博士研究生, 主要从事多媒体信号处理、文语转换、基于内容的音频表征等研究.