University of Leeds

SCHOOL OF COMPUTING

RESEARCH REPORT SERIES

Report 2001.07

Tracking multiple sports players through occlusion, congestion and scale

by

C J Needham & R D Boyle

June 2001

通过阻塞，拥挤和规模模型来追踪多个运动员

摘要

在大的区域内追踪竞技选手是一个具有挑战性的问题。运动员移动迅速，并且他们的身影有较大的变化。

本文提出了一种多目标追踪的框架，采用了 CONDENSATION 为基础的方法。每个被追踪的运动员都被独立地抽象为一个模型，整个小组样本的取样概率是一个基于每个运动员的匹配得分的函数。这个函数奖励一贯良好的成绩，但惩罚一些很好的和一些非常糟糕的匹配分数。地平面的信息一直都在使用，并且该算法的预测阶段被改进之后，和利用卡尔曼滤波器估计出来的位置进行配合。这有助于把每个运动员的估计位置组合起来，并且通过闭塞模型来帮助追踪运动员的位置。

1 介绍

这项工作的目的是追踪运动（特别是足球）球员的移动，利用室内球场中的单一固定摄像机。这是为了让将要采取的行为建模和运动选手的位置分析等工作得以开展。

这项工作有两个主要的动机。首先，体育科学界对于在比赛进程中，知晓运动场地被运动员的覆盖率，以及运动员的移动速度非常感兴趣。根据这些信息可以更具体地设计训练方案，以满足不同的运动员。更有趣的是，团体运动是复杂的活动，它涉及到很多玩家之间的互动。这种多玩家活动允许我们探索更多的关系和相互作用，这种关系和互动不仅仅是运动员之间的，也包括把团队作为一个整体。

这一领域呈现几个挑战性的方面：运动场地的大小意味着运动记录图像的分辨率在场地的最近处和最远处变化很大；体育比赛是繁忙的领域；体育运动员的形状通常在短期内变化巨大；而且运动员以不同的速度移动，经常突然改变方向，这使得他们的运动很难预测。

2 背景

许多不同的各种目的的追踪器在最近几年内被开发出来。其中第一种，专为个体行人监测而设计，这便是 Leeds People Tracker[1]，它采用轮廓追踪，主动形状模型和卡尔曼滤波器，利用一台摄像机追踪多人。一些系统允许测定身体姿势，以及头部和手的实时追踪，如

Pfinder[16]。 Pfinder 是一个“人寻找者”，它使用的颜色和形状的多类统计模型，创建一个斑点表示被追踪的人。这种方法只在场景中仅有一人时有效，并会产生一个比获得运动选手的位置更详细的模型。McKenna 等近期工作中 [10]，在三个层次的抽象中进行追踪：区域，个人和团体。颜色信息在该系统中大量使用，以帮助应对行人场景中的阴影和去闭塞多义性。

目前，有几种常用的方法用于追踪移动目标，其中包括：Active Shape Models[4]，它是灵活的形状模型，允许在对象姿势、尺度和形状的估计中进行迭代精度提升；卡尔曼滤波器 [15,3]，由于其计算效率和其估计未来状态的能力，已被用在许多追踪应用中；以及 Isard 和 Blake 的较新的 CONDENSATION 方法 [11]，这是一种功能强大的技术，它允许条件概率密度随时间的传播，并已和轮廓追踪一起用来追踪杂乱场景中的对象。在追踪多个目标时，这种方法有先天的缺陷。如果使用多个具有相同的追踪算法的单目标的追踪器，那么两个或两个以上的追踪器将会合并到同一个目标，因为这个目标是它们的模型最适合的。近日，MacCormick 和 Blake 推出一种概率排除原则 [8]，来配合 CONDENSATION 追踪，力图解决这个问题。

足球相关工作的灵感来自几个不同的雄心，包括标注，动作识别，比赛重建和比赛评估。Intille 和 Bobick 使用封闭的世界 [6] 视频标注美式足球的画面，一些后续的工作就此展开，以期能在美式足球运动中识别出运动员的动作 [7]。SoccerMan[2] 是一个（足球）比赛重建系统：各种技术被用于追踪运动员，然后一个拥有运动场纹理以及运动员形体纹理的虚拟 3D 世界就可以形成，它可以从任何虚拟视点观看。Taki 等人 [13] 通过调查在球场上的空间优势，来专注于在足球比赛中评估团队表现。

本工作的目的是产生一个追踪器，它将自动追踪运动员，并确定它们的真实世界的位置，以用于进行定位的行为分析，而不是识别他们是否正在跑动，踢球，或涉及一组发挥。

3 理论方面

3.1 图像视角

图 1 足球运动员的示例镜头

通过繁忙杂乱的场景追踪多个目标仍然是一个具有挑战性的问题。图 1 显示出了室内 5 人制足球比赛典型场景。运动员所有的动作被限制在场地中，然而图像透视图突出了几个问题。场地的大小（18x32 米）意味着比赛中图像的分辨率在场地最近处和最远处变化很大。分析显示，在一个典型的图像中（例如，图 1，它是在尺寸 320×240 像素），如果在图像平面上的两个垂直相邻的像素被投影到所述地平面，则在图像的最接近部分的像素相距 3 厘米，而那些在遥远的目标，嘴的大小都超过了 45 厘米。在图像中场地最近部分的区域中，地面的 3 米覆盖了 72 像素，而在场地最远部分，同样的距离在图像中只覆盖了 8 像素。

这强调了考虑图像中景深信息的重要性。在追踪过程中使用地平面坐标变得重要，其中考虑了一个球员在运动过程中物理上能够覆盖的地面范围。在图像中，像素对应的距离变化是非常大的。运动员在地平面的位置信息，能够辅助解决闭塞问题，尤其是从透视图的角度，因为当球员们相距一米以上时，一个球员往往会遮挡其他玩家的一部分。

运动员主要有兴趣的特点是脚的位置，这就是用脚来代替球员的质心的原因，并且在将来的球员建模时，这个位置是我们希望能最大精度确定的。当从图像计算出他们的位置时，假设运动员的脚与地面是接触的。

3.2 图像分割

从视频进行图像分割已经进行了许多尝试，使用背景差分，自适应背景减除[10]，和颜色空间模型[14]。

维护一个临时的背景模型然后进行背景减除已被证明是一种从场景中提取移动物体的快速、高效的方法[1]。这种方法在从较为空旷的场景中提取移动的物体效果最好，但体育活动不属于这一类。体育运动员总是在场地中，并经常（尤其是在像网球这样的运动中）有一些战术位置，他们只在这些位置站立很短的时间，这就可以通过动态背景维护来归入背景模型中了。如果静态背景模型被用来解决这个问题，它可能不容许变化的照明条件或小型照相机的动作。

通过为前景和背景建立现有色彩空间模型[14]，可以进行快速图像分割，具体来说就是对运动员和非运动员建立模型。该方法对于小型相机的抖动和静止的物体非常稳定。

在本工作中，HSI 空间前景模型，从追踪开始之前预先标记为前景的一些图片的像素采样中离线构建。使用 HSI 空间，是因为在 HIS 空间中，前景和背景簇之间的间隔大于其他可能的空间，例如 RGB 或色度值空间。背景也进行同样的处理。对于图像中的每个像素，作为前景的概率按以下公式计算：

$$p(fore) = \frac{d_b}{d_f + d_b} \quad (1)$$

其中d_f和d_b是从各聚类中心的像素的马氏距离。这将创建一个噪点图像，其中运动员的区域被分割开来，尤其是运动员的腿。分割可以通过使用以下概率松弛公式来改进：

$$p(fore) = p(fore) + \delta \quad \text{如果相邻像素的中值 } > 0.5 \quad (2)$$

$$p(fore) = p(fore) - \delta \quad \text{否则} \quad (3)$$

选择δ使得一个像素通过合适次数松弛的应用之后，可以从前景变为背景（或者反之亦然）。应用 3 次概率松弛，取δ = 0.2 可以取得较好的前景分割效果。

3.3 形状模型

图 2 足球动员的形状变化

通常在追踪应用中要识别的对象是在本质上相似的。例如工业检测电阻[4]，行人在停车场[1]，以及锅炉房中的鸡[12]。图 2 示出分离出足球运动员的轮廓的变化，这提出了关于追踪这些形状最佳办法的问题。轮廓模型，如 PDMs[5] 或依赖于抽取出来的形状轮廓点集的样条模型。对于相似的形状，这些方法聚类良好，而且 PCA 可以用来减少维数，从而识别形状的主要特征或特性。单一形状模型看起来不适合建模足球运动员的轮廓。Magee[9] 使用了三个形状模型来代表在追踪奶牛时，不同配置下的牛腿。类似的方法可以应用在这里，使用一定量的模型来代表不同情形下的运动员：当他们双腿并拢站着时；当他们双腿张开站着时；当他们跑动时，形成一个对角线的形状；或者当他们张开双臂时。这种复杂程度对于这种应用来说可能太大。

这里采用的方法就是，为每个轮廓适应边界，并评估轮廓与图像数据的符合程度。如果运动员身体姿势以及方向的信息是我们的目标，使用更复杂的模型也是值得的。追踪的运动员的目的是为分析他们的运动和位置相关的行为，因此最重要的特点是他们的脚。

4 多目标追踪

4.1 结构

一个多对象以 CONDENSATION 为基础的方法被采用，而不是使用多个单目标追踪器。这种多目标追踪为算法的结构增加了一个额外等级。在这里，一个样本代表一个运动员的一个实例，一个样本集表示样本（被追踪的运动员）的实例的集合，而一个超级样本集表示样本集的集合。

运动员的脚与地面的接触点，是我们希望最精确确定的。图像坐标点(u, v)可以用来表示

运动员与地板的接触点；校准图像平面可以使得图像点能被投射到地平面（世界）坐标系中。在本工作中，地平面坐标(x, y)以及由此计算出来的图像位置在整个计算过程中被使用。

要计算一个运动员的边界，首先单一的世界坐标点被投影到图像平面的点(u, v)上；然后形成一个宽为w高为h的边界框；假定该点是边界框的基部的中点。建立的时候，一个标识号，id，被包括进来，用来确定运动员属于哪条轨迹。因此，每个运动员可以被表示为：

$$\mathrm{x} = (x, y, h, w, id) \tag{4}$$

假设X_t^i表示 t时刻一个样本的实例。可以构成一个样本集S_t^j，它由每个不同的被追踪目标的实例组成，以及相应的采样概率π_t^j。

$$S_t^j = (X_t^1, X_t^2, \dots, X_t^{n_j}, \pi_t^j) \tag{5}$$

这里n_j表示样本集S_t^j中包含目标的数量。

超级样本集$S_t = (S_t^1, S_t^2, \dots, S_t^N)$被创建，用于保存每一个样本集，其中$N$是一个在 CONDENSATION 算法中预先定义的样本数。

4.2 传播

超级样本集中的样本集按照常用的方法进行传播，亦即N个样本集每一步按照$p(s_t \mid \zeta_t)$概率进行传播，其中ζ_t为图像中前景概率的数据，亦即，s_t'根据 $p(s_t \mid \zeta_t)$随机绘制。然后，s_{t+1}'根据 $p(s_{t+1} \mid s_t = s_t')$，并且$\pi_{t+1}$根据 $p(\zeta_{t+1} \mid s_{t+1} = s_{t+1}')$计算。

通过一个评估边界与目标匹配程度的匹配度函数，重新调整每个样本集的权重，来重新计算概率。如果一个样本集中每个运动员样本的匹配程度相近，则整个样本集的匹配得分增加（奖励）。如果一个或多个样本匹配程度差，则整个样本集的匹配得分减少（惩罚）。这样做的目的，是帮助样本集的传播与n_j个对象能最好整体配合，而不是那些其中一个或多个对象匹配的非常好，当有一些根本不匹配。具有最高采样概率的样本集被用作“最佳”样本集用于表示运动员。

4.3 预测

从${s_{t+1}}^j \in S_{t-1}$中预测每个样本集${s_t}^j \in S_t$使用的模型为：

$$\begin{aligned} {x_t}^i &= {x_{t-1}}^i + \varepsilon_x & {h_t}^i &= {h_{t-1}}^i + \varepsilon_h \\ {y_t}^i &= {y_{t-1}}^i + \varepsilon_y & {w_t}^i &= {w_{t-1}}^i + \varepsilon_w \end{aligned} \tag{6}$$

其中$i = \{1, \ldots, n_j\}$，而且ε_x和ε_y服从$\mathcal{N}(0, \sigma_1)$分布，σ_1通常在 100 毫米左右，这是考虑到在地平面中，运动员可能运动的最大距离将在$3\sigma_1$的数量级内（每 1/25 秒内 300 毫米）。这允许追踪以 7.5m/s 速度移动的运动员。运动员的速度信息可以这一阶段获得，然而运动的本质经常涉及到运动员进行迅速、突然的方向改变。

由于运动员在一些情形下形状会快速发生变化，例如张开双臂以引起注意，或者跑动的时候跨大步，边界框的宽度和高度必须能够快速响应这些变化，因此在宽度和高度中引入服从$\mathcal{N}(0, \sigma_2)$分布的ε_h和ε_w的高斯噪音，其中$\sigma_2 = 2$像素允许这样的变化。

这样做具有一个缺点，因为样本集中的样本可能不再各自对应于不同的选手，例如当一个样本锁定了另一个非常接近的却早已被追踪的目标时。

5 用卡尔曼滤波器改善

改变 CONDENSATION 算法的预测步骤，能够防止与表示同一目标的样本偏离的样本出现。运动员在地平面上的位置可根据之前的状态在下一步中被预测。这里，对每个运动员使用n_j卡尔曼滤波器。它们根据“最佳”样本集中运动员位置（样本）的观测值进行更新。

使用卡尔曼滤波器是因为它解决了在下一离散时间步骤中，估计运动员的位置$x_t = (x, y) \in \Re^2$ 的问题。一个简单的线性随机差分方程控制这个过程：

$$x_t = x_{t-1} + w_{t-1} \tag{7}$$

其中尺度$z \in \Re^2$，且和x直接相关：

$$z_t = x_k + v_{t-1} \tag{8}$$

独立的随机变量w_t和v_t代表处理过程和测量噪声，并服从正态概率分布：

$$p(w) \sim \mathcal{N}(0, \mathcal{Q}) \qquad p(v) \sim \mathcal{N}(0, \mathcal{R}) \tag{9}$$

目前，使用常数$\mathcal{Q}$（处理过程的噪声协方差）和$\mathcal{R}$（测量噪声协方差）。然而，在将来，这些可被用于评估所提出的估计值的确定性，这将改善在从卡尔曼滤波器估计的运动员位置的“可信度”，相对于从图像解析闭塞时z的观察值。

在每一步中，对于每个运动员位置的一个卡尔曼估计$\hat{\mathrm{x}}_t = (\hat{\mathrm{x}}_t, \hat{\mathrm{y}}_t)$被计算，并且从每个样本集$s^j{}_{t-1} \in S_{t-1}$中预测样本集$s^j{}_t \in S_t$，使用如下公式：

$$\begin{aligned} x_t{}^i &= (\hat{\mathrm{x}}_t + x_{t-1}{}^i)/2 + \varepsilon_x & h_t{}^i &= h_{t-1}{}^i + \varepsilon_h \\ y_t{}^i &= (\hat{\mathrm{y}}_t + y_{t-1}{}^i)/2 + \varepsilon_y & w_t{}^i &= w_{t-1}{}^i + \varepsilon_w \end{aligned} \tag{10}$$

其中$i = \{1, \ldots, n_j\}$，而且最佳样本集中的运动员观测位置z_t被用于更新每个离散的卡尔曼过滤器。这中效果可以根据 CONDENSATION 算法把对应于同一个运动员的样本聚合起来，因为每个样本是对着该运动员的预测值$\hat{\mathbf{x}}_t$绘制的。这可以防止对应于同一个运动员的样本被分裂成两个或多个基团，这可能允许运动员的“最好”样本在组之间跳转，或锁定到不同的运动员。

6 评估和结果

考虑第 3.1 节中的影响，有可能在地平面的有限区域内，地平面的位置的差异的比较是有效的，因为在图像的部分区域内，相邻像素几乎相隔半米。同样，当考虑到不对称形状时，假设玩家的位置是在边界框基部的中点可能是无效的，例如，当一个玩家向一边倾倒时。然而，在这里，假定这些是足够可用有效的。

为了评价追踪，运动员真实地平面上的位置必须被确定。一个序列被独立地手工标记 4 次，并且对和其他轨迹一起的结果轨迹进行分析。图 3（a）显示了在 835 帧中单个足球运动员的轨迹。图 3（b）示出的各轨迹之间欧几里得差的分布，其计算方法为计算每个时间点中两个运动员的距离。对四个手工追踪轨迹的六个成对排列分析显示，位置之间平均的距离是 312.2 毫米，标准偏差为 239.7 毫米，并在 200 和 300 毫米的模式中。因此将四个手工标记的轨迹的平均值作为运动员的“真实”轨迹是合理的，并将之与自动追踪的轨迹进行比较。

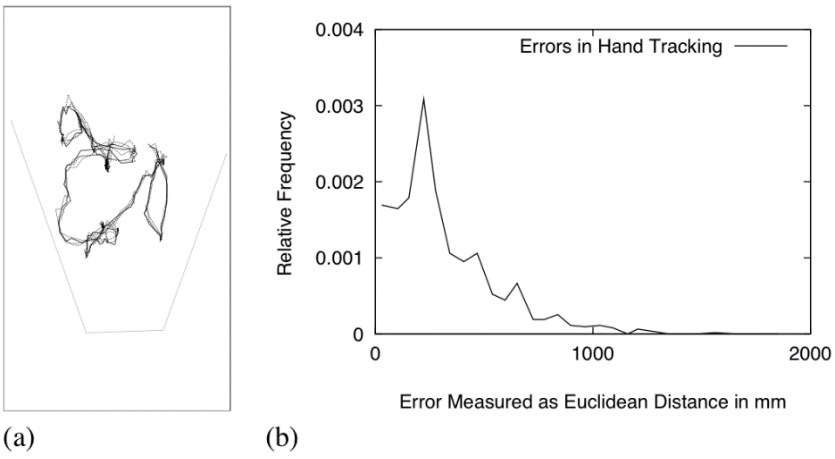
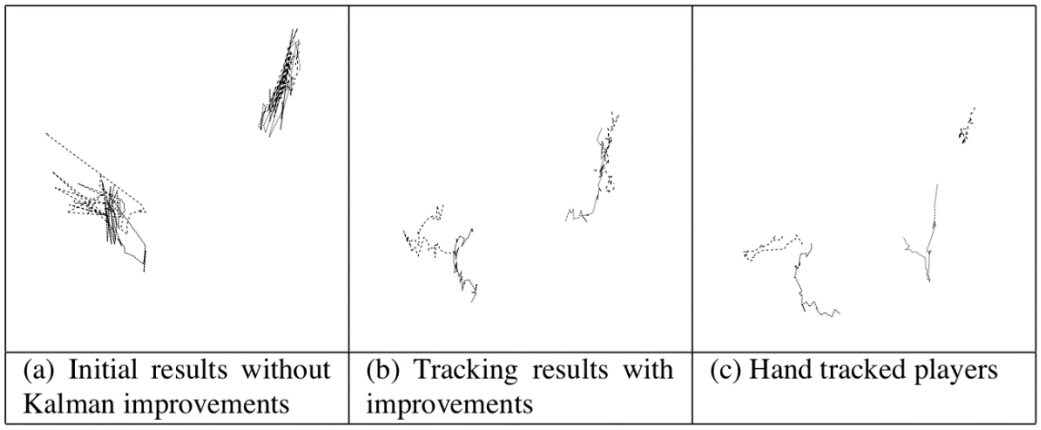


图 3　(a)从 835 帧中每 5 帧一次采样，对同一个足球运动员绘制的四条独立的手工追踪轨迹　(b)在这四条手工追踪轨迹中的 6 对组合

之间的欧几里得距离

对于运动员的行为建模，运动员位置的零误差将是理想的，但鉴于人体机能的变化，高达 0.5 米的误差在手工追踪的数据中可以被认为是可接受的。据预计，在真实位置附近一米的数据，在行为分析中都是可用的，所以如果轨迹的位置在手工追踪平均位置附近一米，都将被视为可接受的。



(a) Initial results without Kalman improvements | (b) Tracking results with improvements | (c) Hand tracked players

图 4　足球运动员在 40 帧中轨迹的比较

对室内 5 人制足球进行一个短序列追踪，四名球员被追踪到。首先，使用的是在第 4 节描述的多目标 CONDENSATION 追踪法，用 N = 1000 个样本，这导致球员的位置被允许在附近跳跃，因为每个球员的多个假设位置被传播开来了。我们观察到，帧到帧之间，样品在多个目标之间切换，而不是始终锁定到一个特定的目标。和手工追踪轨迹相比较，这个不完美的系统显示出了 2.5 米的平均误差，如图 5（a）所示：

进行了在第 5 节中详细描述的改进之后，再次进行追踪，这一次，样本更好地锁定到了四名球员，没有在球员之间切换，也没有多个样本追踪同一名球员。这将位置的平均误差减小到了 1.16 米，并且形状的平均误差低于 400 毫米。图 5（a）示出了误差距离，并突出了新追踪系统中的改进。图 5（b）显示了轨迹的噪音非常小，而且在追踪结束之后使用更多的过滤器平滑轨迹，可以产生更好的结果。图 6 显示了边界框可以标示出被追踪的运动员。

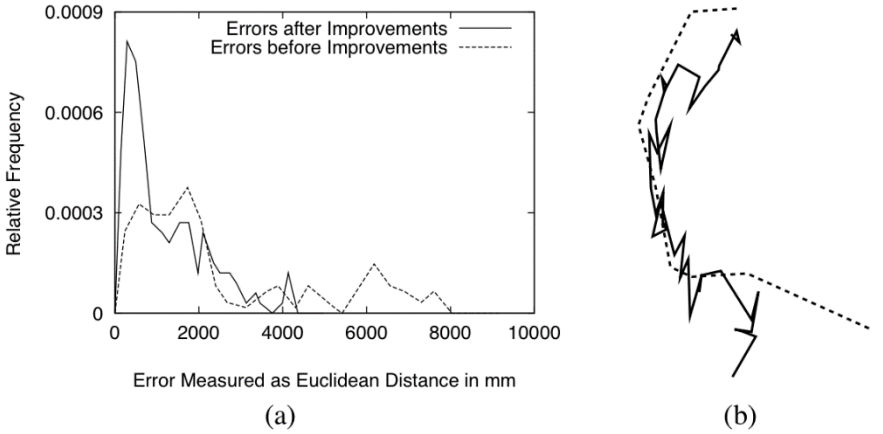


图 5　(a)和手工追踪的序列相比，改进之后，在 40 帧中被追踪足球运动员在地平面中欧几里得距离误差的减少。　(b)足球运动员轨迹的对比。实线表示自动追踪的轨迹，虚线表示手工追踪的轨迹。

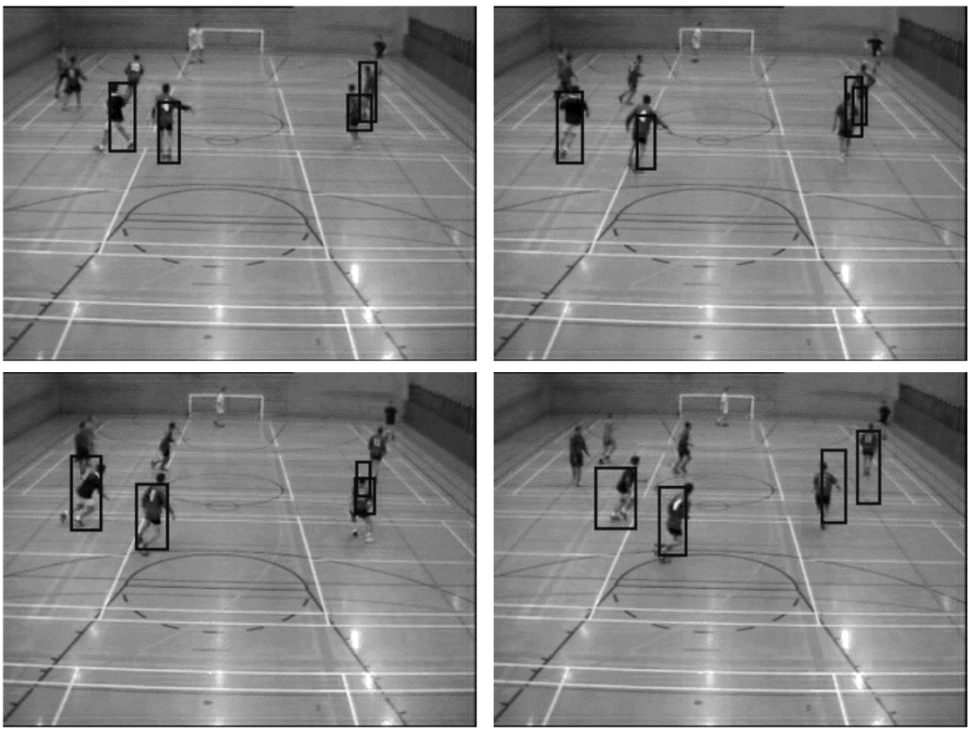


图 6　追踪足球运动员，第 30，40，50，60 帧

7 总结

这项工作提出了一个多目标追踪的新颖框架。最初的方案中 28%的追踪是可用的。经过改进后，56%的轨迹和手工标记的轨迹误差在一米之内，因此可用于行为建模。所描述的边界框方法中，对于脚部的误差并不好，但运动员的追踪效果很好。

今后的工作中会引入更复杂的形状模型，以及对运动员的位置行为分析。

致谢

笔者要感谢大学体育科学系对于获取录像的支持，以及一个 EPSRC 博士助学金奖励的财政支持。

参考文献

[1] A. M. Baumberg. *Learning Deformable Models for Tracking Human Motion*. PhD thesis, School of Computer Studies, University of Leeds, 1995.

[2] T. Bebie and H. Bieri. SoccerMan - reconstructing soccer games from video sequences. In *Proc. of the Int. Conf. on Image Processing*, pages 898–902, 1998.

[3] C.K.ChuiandG.Chen.*KalmanFilteringwithReal-TimeApplications*.Springer, 1999.

[4] T. F. Cootes and C. J. Taylor. Active shape models - ‘smart snakes’. In *Proc. British Machine Vision Conference*, 1992.

[5] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Training models of shape from sets of examples. In *Proc. British Machine Vision Conference*, pages 9–18, 1992.

[6] S. S. Intille and A. F. Bobick. Visual tracking using closed-worlds. In *Proc. Int. Conf. on Computer Vision*, 1995.

[7] S. S. Intille and A. F. Bobick. A framework for recognizing multi-agent action from visual evidence. In *Proc. of the Nat. Conf. on Artificial Intelligence*, pages 518–525, 1999.

[8] J. MacCormick and A. Blake. A probabilistic exclusion principle for tracking multiple objects. In *Proc. Int. Conf. on Computer Vision*, pages 572–578, 1999.

[9] D. R. Magee and R. D. Boyle. Building class sensitive models for tracking application. In *Proc. British Machine Vision Conference*, pages 594–603, 1999.

[10] S. J. McKenna, S. Jabri, Z. Duric, and H. Wechsler. Tracking interacting people. In *Proc. Fourth*

IEEE Int. Conf. Automatic Face and Gesture Recognition, pages 348–353, 2000.

[11] M.Isard and A. Blake. Contour tracking by shochastic propagation of conditional density. In *Proc. European Conf. Computer Vision*, pages 343–356, 1996.

[12] D. M. Sergeant, R. D. Boyle, and J. M. Forbes. Computer visual tracking of poultry. *Computers and Electronics in Agriculture*, 21(1):1–18, 1998.

[13] T. Taki, J. Hasegawa, and T. Fukumura. Development of motion analysis system for quantitative evaluation of teamwork in soccer games. In *Proc. Int. Conf. Image Processing*, 1996.

[14] N. Vandenbroucke, L. Macaire, and J. G. Postaire. Color pixels classification in an hybrid color space. In *Int. Conf. on Image Processing*, pages 176–180, 1998.

[15] G. Welch and G. Bishop. An introduction to the Kalman filter. Technical Report TR 95-041, University of North Carolina at Chapel Hill, 1995.

[16] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: real-time track- ing of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.