

单摄像头实时定位足球

Dmitrii Vorobev¹[0009-0005-2700-1243], Artem Prosvetov¹[0000-0001-7997-0920],
and Karim Elhadji Daou²[0000-0003-4677-7571]

¹ Yandex, Moscow, Russia

² Yandex, Almaty, Kazakhstan

Abstract. 我们提出了一种计算效率高的方法，用于从单个广播摄像机实时重建三维足球轨迹。与之前的工作相比，我们的方法引入了一种具有 W 个离散模式的多模式状态模型，在显著加速优化的同时保持厘米级精度——即使在严重遮挡、运动模糊和复杂背景的情况下。该系统在标准 CPU 上运行，实现了适合直播环境的低延迟。在一个 6K 分辨率的俄罗斯超级联赛比赛专有数据集上的广泛评估表明，其性能可与多摄像系统媲美，而无需专门或昂贵的基础设施。该工作为职业足球环境中提供了一种实用的、可访问且准确的三维足球追踪方法。

Keywords: Sports technology · Real-time localization · Football analytics.

1 介绍

实时的三维足球定位正在成为现代表现分析、裁判判罚和体育转播的一个重要工具。其核心过程包括在每个视频帧中检测球，从校准的摄像机视角（或深度先验）推断其相对于全球球场参考的 XYZ 坐标，并实时过滤这些估计以实现厘米级精度——即使在存在遮挡或运动模糊的情况下。这种精确性使得高级分析成为可能，包括预期进球、射门特征和传球锐度等指标，同时支持需要客观证据的重要技术，如门线裁决和 VAR。精确的球轨迹还允许转播商添加虚拟越位线、动态热图和其他增强现实图形，为观众提供更深入的团队战术见解。而教练则利用这些数据来量化球速和空间压力，将比赛表现与分析反馈结合起来。强大的跟踪技术克服了视频中长期存在的遮挡和模糊问题，使自动化在不同的体育场和摄像机设置中成为可能。最终，精确的球跟踪推动了更丰富的比赛分析和内容创建，为观众带来更具吸引力的体验，使其成为当代足球技术的基石。

然而，挑战依然存在。在一个 6K 视频帧中，球仅覆盖面积的不到 0.15%，并且常常被复杂的背景所掩盖。大约一半的时间，球会被球员、球门柱或裁判完全挡住，其运动轨迹很少遵循简单的抛物线——常受到偏斜、湿滑草皮和误击的影响。尽管多相机设置和全局优化可以提供高精度的结果，但它们需要昂贵的基础设施和专业人员，每个体育场的成本高达数十万美元。相比之下，主要的广播摄像头的画面是普遍可用的且无需额外成本，但从单个视点提取深度信息存在根本性的挑战，使得直接三角测量成为不可能。

我们的目标是开发一种高效的实时方法，从单个广播视角重建球的 3D 轨迹，其准确性可以与多摄像头系统相媲美。我们的解决方案完全在标准 CPU 上运行，并在不到两秒的时间内生成结果，以满足现场直播的低延迟要求。

单摄像机轨迹重建与多摄像机情况有本质上的不同。仅靠一个摄像机角度，实际上无法仅通过图像平面坐标来确定球的真实世界三维位置，因为摄像机到球的距离是未知的（见图 1）。虽然早期的一些尝试 [8] 已经处理过这个问题，但精度仍然不足。

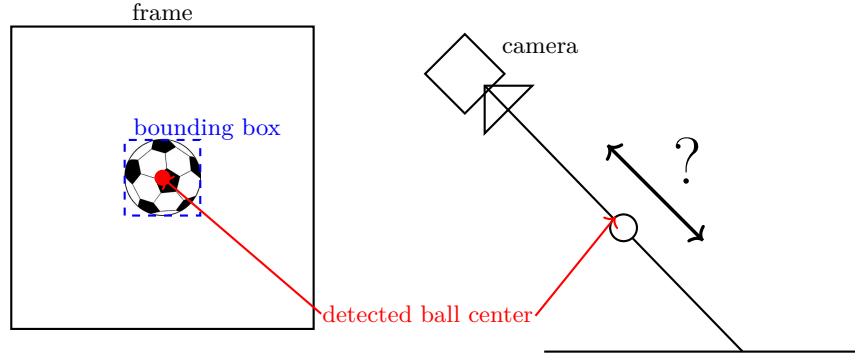


Fig. 1. 仅通过单个摄像机帧确定的球位置的不确定性。

为了解决这些挑战，我们开发了一种多模式状态模型，该模型区分自由飞行、球员持球、踢球后缓冲和出界等阶段。该方法应用有限的离散化和固定宽度的光束搜索，确保每个过滤步骤的线性时间复杂度。我们引入了一个可控的延迟参数 L ，该参数允许在准确性和延迟之间进行动态权衡。我们在一组本公司内部的数据集上验证了我们的解决方案，该数据集来自俄罗斯超级联赛比赛（2024/25 赛季，6K 分辨率），显示出实时性能（单个 Intel-Xeon-CPU-E5-2660 核心上超过 50 fps）以及当 $L = 50$ 帧时达到与离线优化相当的准确性。

总之，我们论文的主要贡献是：一种多模式状态模型，一种线性复杂度优化算法，以及在真实世界数据集上的实证验证。

该工作提供了一个实用的框架，用于在足球直播中仅使用单个标准摄像机监控，无需昂贵的设备，即可实现高精度的 3D 球体追踪。

2 相关工作

现有的三维球体定位方法可以根据两个主要方面进行分类：(i) 使用摄像机的数量和 (ii) 操作的时间模式（即在线与离线处理）。表格 1 总结了与我们的工作相关的每个类别中的关键出版物。

单摄像机离线方法。 早期解决方案通过拟合抛物线到观察到的二维轨迹来估计深度 [1]。Ribnick 等人 [5] 详细展开了这个想法，而 Shen 等人 [6] 则引入了物理（弹道）约束的样条插值来跟踪乒乓球。在他们的方法中，检测跨步的起始和结束位置，并假设球在桌子上的投影进行直线运动来插值轨迹——这个想法也应用于足球，与不同的离散球状态一起，由 Strachan 等人 [4] 提出。Metzler 等人 [3] 基于运动历史图像生成微轨迹，通过物理特性和相机校准来重建三维轨迹。

Table 1. 文献分类法

Method	Cameras	Time	Type
Kim 1998 [1]	1	offline	parabola
Silva 2011 [7]	1	online	EKF+ML
Shen 2015 [6]	1	offline	spline+ballistics
Van Zandycke 2022 [8]	1	online	DL-diameter
Ribnick 2009 [5]	1	offline	fit trajectory
Metzler 2013 [3]	1	offline	fit trajectory
Ren 2008 [4]	4	near-time	EKF
Maksai 2016 [2]	8	offline	MIP

单摄像头在线方法。 Silva 等人 [7] 摒弃了离线优化窗口，使用扩展卡尔曼滤波器（EKF）和短序列上的最大似然估计。然而，他们的方法在长时间的球体遮挡下失效。

最近，利用卷积神经网络的方法已经出现。例如，Van Zandycke 等人 [8] 使用单帧网络直接估计球体直径（从而测算深度），虽然实现了高速，但由于没有进行物理建模，因而缺乏速度估计或事件检测。

多摄像头、近实时和离线方法。 Ren 等人提出了一种使用四个静态摄像头的近实时算法，延迟为 $\sim 1 - 2$ 秒。Maksai 等人通过对 500 帧的全局 MIP 优化实现了高精度，但代价是解决时间更长（数十秒）。这两项工作都使用潜在变量分解了球的离散状态。

所提方法的定位。 我们的算法结合了单摄像头设置的经济性和简单性，以及在线操作的低延迟（最低至单帧），通过离散事件模型得到了增强。通过这样做，我们弥合了快速但有限的单摄像头 EKF 方法与更准确但代价高昂且复杂的多摄像头利用潜在状态模型的算法之间的差距。

3 问题表述

我们研究如何从广播视频中重建足球的三维轨迹。该问题的输入是球和球员在场上位置的检测结果，外加每个时间步的摄像机标定参数。多个因素使这个任务变得复杂：球经常被球员遮挡，导致有些帧中球不可见；此外，与球员和环境的交互以及突然的轨迹变化，使得精确的重建具有挑战性。

为了克服这些困难，我们明确地对球、球员和环境之间的交互进行建模。在数学上，这个问题被表述为在给定观测数据的条件下，最大化球轨迹的后验概率。通过模拟可能的球状态并修剪概率低的状态，在线推断解决方案。

3.1 输入数据和预测

对于时间步 t 的每一帧，算法接收以下输入：屏幕上球的位置 $U_t = \{u_t^{(i)} \in \mathbb{R}^2\}$ ，场地上球员的位置 $G_t = \{g_t^{(i)} \in \mathbb{R}^2\}$ ，以及相机校准参数 K_t, R_t, T_t ，其中 K

是固有相机矩阵, R 是旋转矩阵, T 是平移向量。作为输出, 对于每一帧 t , 我们的算法返回在空间中直到时间 t 的球的最可能轨迹。

表示观察到的变量 $\mathcal{Z}_{1:T} = \{U_{1:t}, G_{1:t}, K, R_{1:t}, T_{1:t}\}$

$$X_{1:T}^* = \arg \max_{X_{1:T}} p(X_{1:T} | \mathcal{Z}_{1:T}). \quad (1)$$

4 在线解决方案

我们的方法实现为一个递归滤波器, 具有混合的连续-离散状态和轨迹空间内的有限搜索宽度。在每个帧中, 我们生成有限的假设集, 并为每个假设分配一个对数似然分数。仅保留最有可能的 K 个假设, 其余全部舍弃。该方法确保计算复杂性随检测数量线性增长。

我们将 (1) 重新表述为带有隐藏变量 H_t 的马尔可夫模型:

$$p(X_{1:t}, H_{1:t} | \mathcal{Z}_{1:t}) := p(X_t, H_t | X_{t-1}, H_{t-1}, \mathcal{Z}_t) \cdot p(X_{1:t-1}, H_{1:t-1} | \mathcal{Z}_{1:t-1})$$

4.1 潜在变量

滤波器的潜状态 H_t 包括一个离散模式 $s_t \in S = \{\mathcal{J}, \mathcal{P}, \mathcal{W}, \mathcal{O}\}$, 表示跳跃、持球、等待和出界状态, 以及其他参数 (详情见 Sec. 5.1)。

4.2 对数似然和剪枝

我们将对序列 (X_t, H_t) 记作 $\mathcal{X}_t$ 并将似然乘子分解为三个部分:

$$p_s(s_t + 1 | s_t) \cdot p_x(X_{t+1} | X_t, s_{t+1}) \cdot p_z(\mathcal{Z}_{t+1} | X_{t+1}, s_{t+1})$$

。在这里, p_s 是一个由历史数据预先计算的硬编码转移矩阵。 p_x 和 p_z 的计算方法在第 6 节详细说明。

在帧 t 修剪后, 我们保留了一组假设 $\{\mathcal{X}_t^{(n)}\}_{n=1}^{N_t}$ 及其对应的累积对数权重 $\ell_t^{(n)}$ 。每个假设产生一组后代 $\{\mathcal{X}_{t+1}^{(n,m)}\}$: 索引 n 指示父项, 而 m 在帧 $t+1$ 中索引特定的后代。过渡的对数似然增量 $\mathcal{X}_t^{(n)} \rightarrow \mathcal{X}_{t+1}^{(n,m)}$ 由以下给出:

$$\begin{aligned} \Delta \ell_{t+1}^{(n,m)} = & \underbrace{\log p_s(s_{t+1}^{(n,m)} | s_t^{(n)}) + \log p_x(X_{t+1}^{(n,m)} | X_t^{(n)}, s_{t+1}^{(n,m)})}_{\text{dynamics}} + \\ & \underbrace{\log p_z(\mathcal{Z}_{t+1} | X_{t+1}^{(n,m)}, s_{t+1}^{(n,m)})}_{\text{observation}} \end{aligned}$$

每个后代的总权重是递归计算的为 $\ell_{t+1}^{(n,m)} = \ell_t^{(n)} + \Delta \ell_{t+1}^{(n,m)}$ 。在评估所有后代之后, 合并结果集, 按 $\ell_{t+1}^{(n,m)}$ 降序排序, 并保留前 K 个假设, 舍弃其余的。

4.3 具有大小为 L 的在线缓冲区的链提取

在每一帧中，上述算法产生了一组候选假设 $\{\mathcal{X}_t^{(n)}\}$ 。我们选择对数似然值最高的假设 $\ell_t^{(n)}$ ，然后递归回溯其祖先链以恢复轨迹 $X_{1:t}$ 。

然而，仅轨迹中的一部分直到帧 $t-L$ 会被立即输出；最新的 L 估计保持软状态，当接收到新的观测时可以被修正。此缓冲机制允许对算法延迟进行动态控制，并在不需要立即逐帧估计时提高预测准确性。

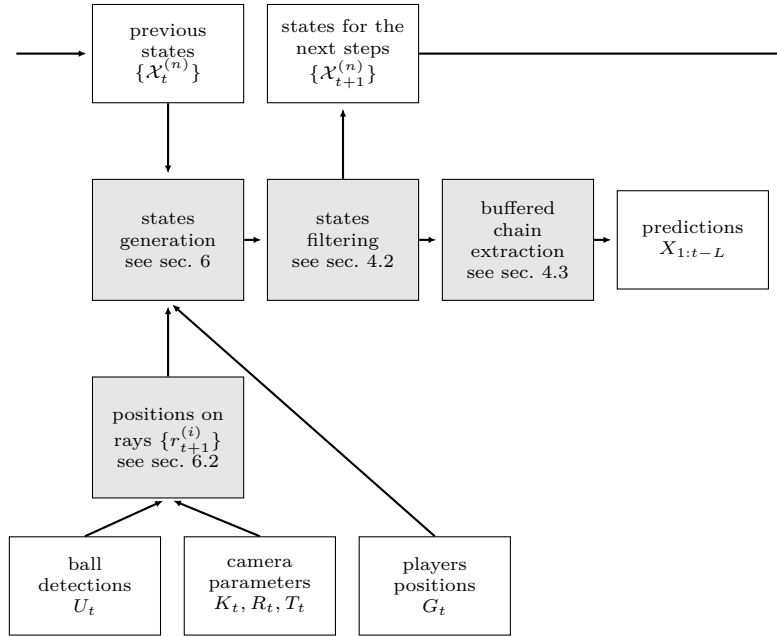


Fig. 2. 所提出流程的方案。

5 鲍尔州立大学模式

5.1 球运动的离散模式

为了模拟游戏事件的多样性，我们定义了一组有限的离散球状态 $S = \{\mathcal{J}, \mathcal{P}, \mathcal{W}, \mathcal{O}\}$ 。对于每一帧 t ，离散变量 s_t ，包含在潜在状态 H_t 中，指定了适用的运动学模型。

在跳跃模式下，球没有接触任何玩家，仅在重力的作用下移动，可能在场地上弹跳。在这种状态下，玩家之间没有交互。除了位置之外， $\mathcal{J}$ 状态还存储 v_t （当前球速）和 Σ_t （位置方差）。 $\mathcal{J}$ 状态可能是可见的或被遮挡（阻挡或未检测）。生成 $\mathcal{J}$ 状态的算法在第 6 节中详细描述。

在控球模式下，球的中心与球员的坐标对齐，其独立运动被抑制。球员的位置生成对应的 $\mathcal{P}$ 状态，每帧为每位球员生成一个。

W —等待占有后 等待模式是在球离开球员的控制但在视觉检测到之前的一个缓冲状态，支持跳跃阶段的稳健初始化。除了位置之外， W 状态还维护 b_t （球被踢时的帧）和 g_t （踢球者的坐标）。

在 $s_t = W$ 时，滤波器只存储踢球时间和二维球员坐标，不生成弹道假设。只有当球首次变得可见时， u_t^i 出现，根据摄像机的视线推断可能的位置，并将状态转换到 J 模式。这种设计确保弹道计算仅在每次出现时运行一次，而不是在每个控球帧上运行，从而显著减少所需假设的数量，保持实时波束宽度达到 $K \approx 1000$ 。

W 态是从前一帧的 W 态或当前的 P 态生成的。

O —走调 越界状态模型描述了球在场外等待重新进入比赛的时间段（例如，界外球或角球）。在这种模式下，不存储额外的潜在变量。 O 状态是从当前帧中被判定为在足球场边界之外的 J 状态生成的。

6 状态转换

图 3 展示了随时间变化的所有可能的球状态之间的转换。有些转换，例如 $O \rightarrow O$ 、 $W \rightarrow W$ 、 $J \rightarrow O$ 、 $P \rightarrow W$ ，可以通过简单继承前一状态的位置来处理。在这些情况下，新状态的位置 X_t 是直接从前一状态复制的。对数似然增量 $\Delta\ell$ 完全依赖于转换概率 p_s 并简化为 $\Delta\ell = \log p_s(s_{t+1} | s_t)$ 。

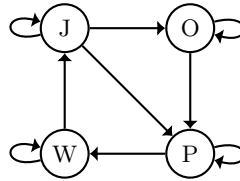


Fig. 3. 状态转换的可能类型。

6.1 any $\rightarrow P$ 跃迁

对于进入玩家持球状态（ P 状态）的转换，似然对数增量 $\Delta\ell$ 另外使用函数 p_x ，该函数只能取值 0 或 1，并由 X_{t+1} 和 X_t 之间的距离决定：

$$p_x(X_{t+1} | X_t, s_{t+1} = P) = \|X_{t+1} - X_t\| < \text{thrld}$$

在这里，thrld 是从历史数据预计算的阈值。因此，如果新位置 X_{t+1} 足够接近先前位置，则 $\Delta\ell = \log p_s(s_{t+1} = P | s_t)$ ；否则， $\Delta\ell = -\infty$ ，有效地消除了不太可能的转换。

6.2 从可见检测生成假设

给定摄像机的校准参数和图像中的球检测，球的三维位置无法唯一确定；相反，可能的位置信息被限制在从摄像机点出发的一个射线上。对于每个检测，我们将这一射线从场地表面到摄像机的位置以 3 厘米的增量进行离散化，并生成候选的 J 状态（见图 4）。我们将这些可见位置记为 $\{r_t^{(i)}\}$ 。

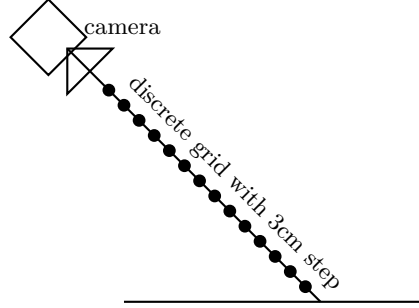


Fig. 4. 为可见 J 态生成的位置

6.3 $\mathcal{J} \rightarrow \mathcal{J}$ 转变

在跳跃模式下，当前 $\mathcal{J}$ 态会生成可见和不可见的 J 态，这取决于是否有检测与之关联。首先，通过在时间步长 dt 上外推前一状态的动态来创建不可见状态，其中 dt 是连续帧之间的时间： $dt = 1/\text{fps}$ 。使用之前的空间位置和速度向量来计算新位置。我们引入外推函数 $j(X_t, v_t, dt)$ ，它产生更新后的位置和速度向量 X'_{t+1}, v_{t+1} 。为了计算可见状态的可能性，我们还需要空间位置方差，这描述了当前位置的不确定性。我们按如下方式更新方差： $\Sigma'_{t+1} = \Sigma_t + dt \cdot \Sigma_v$ ，其中 Σ_v 是从历史数据中得出的。对数似然的增量由 $\Delta \ell = \log p_s(s_{t+1} = \mathcal{J} | s_t = \mathcal{J}) + \log p_z^{\text{invisible}}$ 给出，其中 $p_z^{\text{invisible}}$ 是预先计算的值，用于在没有检测的情况下惩罚状态。

为了生成可见的假设，我们将不可见的候选与可用的检测关联起来。对于每个离散射线位置 $r_{t+1}^{(i)}$ （见第 6.2 节），形成一个新的可见假设： $X_{t+1}^{\text{visible}} := r_{t+1}^{(i)}$ 。似然分量 p_x 被计算为一个均值为 X'_{t+1} 、方差为 Σ'_{t+1} 的高斯分布在 $X_{t+1} = r_{t+1}^{(i)}$ 处的值，即 $p_x = N(r_{t+1}^{(i)} | X'_{t+1}, \Sigma'_{t+1})$ 。这个公式惩罚偏离物理合理性的轨迹。因此，球的轨迹越遵循物理规律，惩罚就越小。因子 p_z 设为 1， Σ_{t+1} 设为 0。

6.4 $\mathcal{W} \rightarrow \mathcal{J}$ 转换

对于光线 $r_{t+1}^{(i)}$ 上的每个可见位置（参见第 6.2 节）和 $\mathcal{W}$ ——一个新的 J 状态被初始化，位置为 $X_{t+1} = r_{t+1}^{(i)}$ 。速度向量通过自由飞行方程计算，使用球的初始位置 X_t 、终点位置 X_{t+1} ，以及球未被检测到的时间（该时间是类型 $\mathcal{W}$ 的隐藏状态的一部分）。乍一看，新的 J 状态的方差 Σ_{t+1} 应该是最小的，因为我

们已经明确地对其当前位置进行了建模；然而，我们赋予它一个预定的值 Σ_g 。虽然新的位置通过检测被精确建模，其起点（确切的踢球位置）则是不确定的——因为只有在踢球瞬间的球员位置被观察到，而用来踢球的身体部位可能有所不同（飞行方向显著地依赖于使用的身体部位）。我们不是引入速度向量的方差，而是将上述不确定性分配给 Σ_{t+1} ，从而减少 J 状态的潜在变量数量。这种方法在数学上是一致的，因为后续的计算仅依赖于位置和速度的组合方差。

这里，对数似然函数的增量仅取决于 $p_s(s_{t+1} = \mathcal{J} \mid s_t = \mathcal{W})$ 。

7 实验

7.1 数据

为了进行定量分析，我们使用了来自 2024/25 赛季俄罗斯超级联赛广播录像中收集的数据集。样本包括 10 场比赛的 3 分钟片段（总共 30 分钟）。源视频以 6144×3240 像素分辨率和 25 帧每秒的速度捕获。对于每一帧，通过使用放置在体育场周围的四个校准摄像机的多视角三角测量获得了球的真实位置。使用控制标记测量的平均三角测量误差为 < 10 厘米。离散模式 $\mathcal{J}, \mathcal{P}, \mathcal{W}, \mathcal{O}$ 由人工标记。使用 RTMDet-t 检测器获取在广播视角中的球的像素坐标。使用相同的检测器获取的球员检测结果，然后使用单应性投影到球场平面上。

7.2 基线方法

为了评估所提出的在线滤波器，我们选择了一组代表不同世代方法的单目算法。

1. Kim 1998 [1] — 用抛物线逼近二维轨迹，并根据投掷参数恢复深度。通过一次非线性最小二乘求解器的运行，对整个片段进行离线操作。
2. Metzler 2013 [3] — 基于运动历史图像技术的三维重建：先在二维中分割球体轨迹，然后拟合物理轨迹；专为静态摄像机设计，离线操作。
3. Ribnick 2009 [5] — 分析性 EKF，用于推导条件以从单个摄像机独特恢复抛射物的深度和速度；不考虑游戏场景或玩家交互。
4. 2016 年 Maksai [2] — 在 500 帧窗口上的多摄像头 MIP 优化；包括作为离线准确度的上限，但在速度和延迟方面不可比较。

7.3 评估指标

定位精度。 我们使用一个阈值度量 $\text{Accuracy}@d$ — 即在位置误差不超过给定阈值 d 的帧的比例。

结果报告为 $d \in \{0.5, 1, 2, 4, 8\} \text{ m}$ 。这一尺度同时提供了“远程精度”（ ≤ 50 厘米，适用于预期目标分析）和粗略定位（ ≤ 8 米），足以在电视广播中进行可视化。

对于关键的离散转折点：踢和跑出音调（OOP） F_1 分数使用以 GT 标签为中心的 ± 12 帧的容差窗口进行计算。

Table 2. 准确性和速度的比较。最佳结果用粗体表示，次佳结果用下划线标出。

Method	Accuracy@ d (m)					F_1 events		fps	Latency
	0.5	1	2	4	8	kick	OOP		
Ours, $L=1$	0.50	0.61	0.65	0.69	0.71	0.59	0.64	53	1
Ours, $L=10$	0.56	0.63	0.66	0.69	0.71	0.61	0.66	53	10
Ours, $L=25$	<u>0.59</u>	0.64	0.67	0.70	0.71	0.71	0.72	53	25
Ours, $L=50$	0.59	0.66	<u>0.67</u>	<u>0.71</u>	<u>0.71</u>	0.74	<u>0.79</u>	53	50
Kim 1998	0.21	0.39	0.52	0.61	0.68	—	—	9	4500 [†]
Metzler 2013	0.34	0.49	0.59	0.64	0.69	—	—	6	4500 [†]
Ribnick 2009	0.30	0.56	0.63	0.67	0.69	—	—	21	4500 [†]
Maksai 2016 (MIP)	0.56	<u>0.64</u>	0.68	0.73	0.77	<u>0.66</u>	0.81	6	500 [†]

[†] For offline methods, latency is equal to the length of the three-minute clip ($3 \text{ min} \times 25 \text{ fps} = 4500 \text{ frames}$); fps is computed as the number of frames divided by the total processing time.

速度和延迟。 吞吐量 (fps) 是每秒处理的帧数。延迟被定义为从帧可用的时刻到算法输出该帧估计值的时刻之间的差异，并在测试集上取平均值 (以帧表示)。对于离线方法，给出处理剪辑的总时间，转换为等效的 fps，并将延迟设置为与剪辑长度相等。为了分析准确性和延迟之间的权衡，如 Sec. 4.2 所述，我们的算法使用可调节的延迟缓冲参数 $0 \leq L \leq 50$ 帧来运行。

7.4 结果分析

表 2 显示，在延迟最小为 $L=1$ 帧的情况下，我们的滤波器在所有精度阈值上都不如离线的 Maksai 2016 (多摄像机 MIP) 方法。然而，将延迟增加到 $L=25$ 几乎缩小了精度差距，而在 $L=50$ 帧时，我们的在线解决方案在某些方面变得可以与 MIP 基线相媲美，甚至超过。

- 在远程精度水平 $\text{Acc}_{0.5}$ 和 Acc_1 ，我们的过滤器分别比 MIP 提高了 +0.03 和 +0.02 (0.59 对 0.56 和 0.66 对 0.64)；
- 对于比赛事件，踢球的 F_1 更高 (0.74 对 0.66)，这对于实时分析射门特别有价值；
- 即使在最高延迟的情况下，我们的方法依然保持实时吞吐率 (50 fps)，而这与显著更慢的 MIP 方法 (6 fps) 形成对比。

我们算法的优势来自于其为足球量身定制的设计：离散模式 $\mathcal{P}, \mathcal{W}, \mathcal{O}$ 明确建模控球、踢球和界外事件， $\mathcal{W}$ 缓冲区推迟了昂贵的弹道更新，直到出现可靠的检测结果为止。Maksai 的通用 MIP 求解器并不区分这些微事件，因此在拥挤或被遮挡的场景中积累了误差。

总之，对于实际延迟 $L \leq 50$ 帧 (在 25 fps 下为 2 秒)，我们的在线滤波器在精度上与离线优化相匹配或超过离线优化，同时实现了计算成本的数量级减少 (推断时间减少了 ~ 10 倍)。

8 结论

在这项工作中，我们提出了一种新颖的多模式状态模型，专为足球转播分析的需求量身定制。通过将过滤器结构化，以在一组有限的离散且可解释的球状态上运行（包括跃起（ $\mathcal{J}$ ）、球员控球（ $\mathcal{P}$ ）、控球后等待（ $\mathcal{W}$ ）以及离场（ $\mathcal{O}$ ）），我们实现了高效的计算和特定领域的建模。该设计使我们能够捕捉到诸如踢球和控球变化等微事件，同时通过引入 $\mathcal{W}$ 缓冲区，高效连接了踢球和首次检测到球之间的时刻。因此，与弹道轨迹假设相关的计算复杂度大大降低。

我们在来自俄罗斯超级联赛的高分辨率广播录像上进行的实验评估表明，在适中的延迟设置（ $L \leq 50$ 帧）下，所提出的在线滤波器在定位和事件检测的准确性方面达到或超过了 Maksai 等人提出的最先进的离线优化方法，但其计算成本仅为约十分之一。值得注意的是，我们的滤波器在关键比赛事件（如踢球）中实现了更优的远程准确度（Acc0.5, Acc1）和更高的 F1，同时保持真正的实时吞吐量（50 fps）。

这些结果强调了将领域知识整合到体育分析的状态估计流程中的好处。所提出的方法能够快速、精准且可解释地分析广播视频，从而促进从现场体育分析到电视可视化的各种应用。在未来的工作中，我们计划丰富所建模的足球事件集，并研究适用于多种广播和分析基础设施的稳健部署策略。

References

1. Kim, T., Seo, Y., Hong, K.S.: Physics-based 3d position analysis of a soccer ball from monocular image sequences. In: Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271). pp. 721–726. IEEE (1998)
2. Maksai, A., Wang, X., Fua, P.: What players do with the ball: A physically constrained interaction modeling. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 972–981 (2016)
3. Metzler, J., Pagel, F.: 3d trajectory reconstruction of the soccer ball for single static camera systems. In: MVA. pp. 121–124 (2013)
4. Ren, J., Orwell, J., Jones, G.A., Xu, M.: Real-time modeling of 3-d soccer ball trajectories from multiple fixed cameras. IEEE Transactions on Circuits and Systems for Video Technology **18**(3), 350–362 (2008)
5. Ribnick, E., Atev, S., Papanikolopoulos, N.P.: Estimating 3d positions and velocities of projectiles from monocular views. IEEE transactions on pattern analysis and machine intelligence **31**(5), 938–944 (2008)
6. Shen, L., Liu, Q., Li, L., Yue, H.: 3d reconstruction of ball trajectory from a single camera in the ball game international. In: Symposium on Computer Science in Sports (IACSS 2015). pp. 33–39 (2015)
7. Silva, H., Dias, A., Almeida, J., Martins, A., Silva, E.: Real-time 3d ball trajectory estimation for robocup middle size league using a single camera. In: RoboCup 2011: Robot Soccer World Cup XV 15. pp. 586–597. Springer (2012)
8. Van Zanduycke, G., De Vleeschouwer, C.: 3d ball localization from a single calibrated image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3472–3480 (2022)