arXiv:2601.19036v1 [cs.GR] 26 Jan 2026

迈向制作就绪的最后一公里：基于视频的捕捉的基于物理的动作优化

Tianxin Tao
ttao@ea.com
SEED 艺电 温哥华，加拿大

Han Liu
haliu@ea.com SEED 艺电 红木城，美国

Hung Yu Ling
hling@ea.com 艺电 温哥华，加拿大

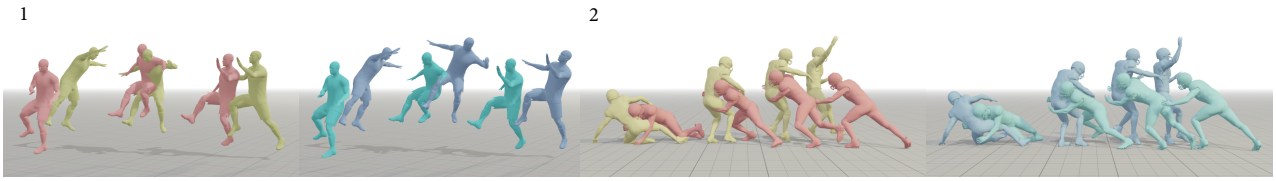


图 1：我们提出了一个框架，可根据不完美的基于视觉的捕捉数据生成物理上合理的动作。图中展示了两个动作片段：1. 头球时的躯干穿透；2. 动作重定向造成的手臂与躯干穿透。黄色和粉色角色：原始动作。青绿色和蓝色角色：优化后的结果。

摘要

高质量动作数据是游戏、电影、扩展现实和机器人学的基础。基于视觉的动作捕捉工具取得了显著进展，能够以较低门槛提供视觉上令人信服的结果，但由于捕捉过程中引入了各种伪影，在最终阶段——即最后一公里——往往仍无法达到物理真实感和制作就绪的要求。本文通过案例研究和专业动画师的反馈总结关键问题，为未来的动作清理研究奠定基础。随后，我们提出一种基于物理的动作优化框架来弥合这一差距，旨在减少耗费大量人力的手动清理工作，并提升视觉质量和物理真实感。该框架支持单角色和多角色序列，也可集成到动画师工作流程中，以便进一步优化，例如通过关键帧编辑实现动作风格化。

ACM 分类概念

• **计算方法** → **动作处理**.

ACM 参考文献格式：
Tianxin Tao、HanLiu 和 Hung Yu Ling。2025。迈向制作就绪的最后一公里：面向基于视频的捕捉的基于物理的动作优化。收录于*2025*年亚洲计算机图形学大会技术交流（*SA Technical Communications '25*），*2025*年*12*月*15*日至*18*日，香港，香港。ACM，纽约，纽约州，美国，4页。
https://doi.org/10.1145/3757376.3771379

1 引言

动作捕捉（动作捕捉）是角色动画的核心。它已从传统的基于标记点和基于惯性传感器的系统发展到近期的基于视觉的技术，大幅减少了设置工作量，并实现了可扩展的动作数据采集。基于视觉的捕捉无论是在学术界 [Kanazawa 等，2019；Sun 等，2023；Ye 等，2023]还是在商业产品中 [DeepMotion 2025；Hawk- eye Innovations 2025；Move.ai 2025；Tracab Technologies 2025],如今都能在多种环境中提供视觉效果出色的结果。然而，由于存在细微但关键的物理不合理伪影，原始的视觉捕获数据仍需人工清理才能达到生产质量。在本文中，我们首先旨在通过收集专业动画师的反馈与见解，系统地识别和分类漂浮、网格相互穿透、幻影接触和脚部滑动等动作伪影，并指出即使在高质量的商业流程中，这些问题仍然存在，从而限制了捕获动作的大规模部署。为弥合这一"最后一公里"差距，我们构建了一个基于物理的动作优化框架，使用深度强化学习（DRL）逐序列运行，在保留原始内容完整性和保真度的同时，修正存在缺陷的动作。该框架同时支持单角色和多角色场景，并引入了伪影感知评估、自适应终止和灵活的奖励塑形等关键创新。

2 相关工作

动作清理主要有两种途径：基于物理的方法和数据驱动方法。数据驱动方法 [Chen 等，2021；Holden 2018]需要大量高质量数据集，而这类数据集并非总是可用，因此我们重点关注基于物理的方法。早期研究通过模拟标记点与角色之间的作用力来提升质量 [Zordan 和 Van Der Horst2003], ，并通过基于采样的控制解决地面穿透等伪影 [Liu 等， 2010]。后续研究将这些方法拓展到了更广泛的场景 [Peng 等，2018b；Yu 等， 2021]。近期，基于深度强化学习的模仿能够改进由 3D 姿势估计器生成的动作 [Yuan 等，2021；Zhang 等，2023b], 、涉及多角色交互的动作 [Huang 等， 2022], 以及重定向后的

本作品采用知识共享署名-非商业性使用 4.0 国际许可协议。*SA Technical Communications'25*，香港，香港© 2025 版权所有者/作者保留版权。ACM
ISBN 979-8-4007-2136-6/2025/12
https://doi.org/10.1145/3757376.3771379

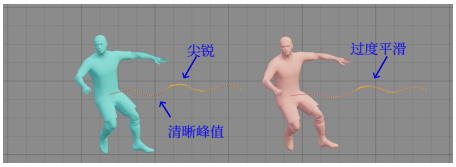


图2：一段尖锐的定格动画，显示原始输入（粉色）中经过过度平滑处理的髋部（根部）轨迹，缺乏重量感。我们的方法生成的精炼动作（青绿色）增添了逼真的根部运动。

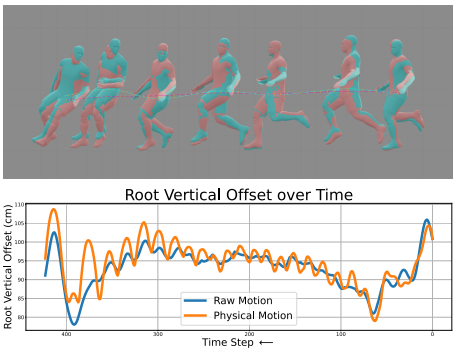


图 3：运动行走过程中被过度平滑的髋部轨迹。该图对比了原始动作（粉色）与我们的物理重建（青绿色）中的髋部高度。

带有接触图的动作 [Zhang等人，2023a]。Ma 等， [2021] 提出了时空界限，作为跟踪奖励的替代方案。

3 种动作伪影

利用物理原理优化动作的一个关键动机在于，许多伪影都源于违反物理定律。物理真实感是高质量动作的必要条件——任何违背重力、摩擦力或碰撞约束等物理原理的动作，本质上都有缺陷。我们重点介绍影响捕获数据质量的常见物理伪影。

漂浮与重量感。 捕获数据中的一个典型问题是缺乏重量感。角色常常看起来像是在滑行或漂浮，尤其是在奔跑或突然停止等动态动作中（见图 2），即使双脚看起来稳稳地踩在地面上也是如此。这通常是由根部轨迹过度平滑造成的，它会削弱垂直运动并降低冲击力，如图 3 所示。专业动画师很容易将过度平滑的根部轨迹识别为低质量动作。基于物理的模拟器通过模拟重力和惯性力，有助于解决这一问题，使角色呈现自然且接地的交互。

穿透。自穿透和角色间穿透在单人和多人动作中都很常见。图4展示了这些伪影，例如手臂与躯干相交或腿部发生碰撞，这是因为姿态估计器未能考虑物理碰撞所涉及的体积。基于物理的仿真器会在每一步自动施加碰撞约束，因此是解决这些问题的理想工具。

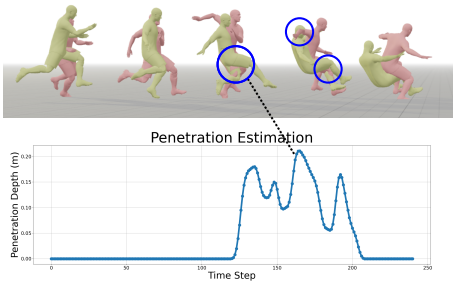


图4：两个角色相互冲撞的动作片段中的穿透深度。穿透深度峰值以蓝色圆圈标出。

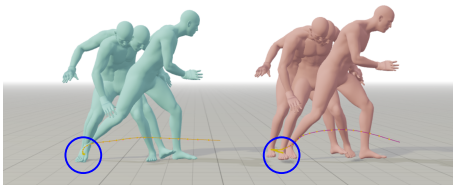


图5：脚部滑动示例。粉色角色出现滑动伪影，而我们方法生成的青绿色版本则保持了正确的脚部接触。蓝色圆圈标记的是一只应保持静止、已着地的右脚。

幻影接触。角色与物体互动时，接触时序总是难以准确捕捉，因而会出现幻影接触。这通常源于物体建模不完整。对于较简单的物体，可以将物体运动跟踪纳入基于物理的仿真，以增强接触真实感，从而解决这一问题。然而，对所有物体进行完整建模仍不切实际。作为替代方案，我们采用动画师熟悉且广泛使用的关键帧技术来指导动作清理。动画师手动添加接触关键帧，系统则自动优化动作，在保持真实感的同时减少伪影。

脚部滑动。 脚部滑动源于对全局轨迹和关节位置（尤其是脚部）的预测误差，因为脚部是较为敏感的末端执行器。在基于视觉的捕捉中，这会导致角色在地面接触期间出现明显的漂移，如图5所示。基于物理的模拟器通过对地面摩擦力进行建模，使脚部保持稳固着地，从而缓解这一问题。

4 系统概述

我们的框架旨在解决前述四种动作伪影。基于物理的动作跟踪能够自然地缓解漂浮和脚部滑动问题；而穿透问题则需要额外处理，因为在不损害运动保真度的前提下解决几何碰撞十分复杂。幻影接触则根据物体交互的具体情况逐一处理。

我们的框架首先评估动作捕捉中伪影的严重程度。根据观察，我们发现，测量穿透程度是判断伪影严重程度的有效指标。

具体而言，我们评估自穿透，以解决肢体穿插等常见问题，这类问题通常发生在单个角色身上。对于涉及多个角色的复杂交互，我们会测量角色之间的穿透程度，以确定所需的修正幅度。随后，我们在标准的基于物理的动作模仿设置上进行改进，类似于 DeepMimic [Peng et al. 2018a]，主要修改了提前终止、奖励设计和动作空间。

4.1 穿透估计

穿透深度直观地衡量物体相交的程度。深度越大，网格穿插问题就越明显、越不真实，也越需要修正。我们使用关节式人形模型，先进行碰撞检测，再估计穿透深度，具体采用 Gilbert-Johnson-Keerthi [Gilbertet al. 1988] 算法和扩展多面体算法 [Van Den Bergen 2001]。

在图4中，我们展示了 ρ_t 一段动画片段中的穿透深度，该片段呈现了两个角色并排相撞的场景。在每个时间步 t，给定所有穿透对后，穿透深度 ρ_t 取所有身体部位对中的最大值。穿透深度可作为伪影存在位置的可靠指标。它能够精准捕捉常见问题，例如腿部穿插、肢体重叠，以及角色交互过程中的穿透。

4.2 动作模仿

状态与动作。状态向量 s_t 包含根节点信息，包括根节点高度、根节点线速度、根节点角速度和根节点朝向；还包含关节级属性，包括关节朝向 q_t 和关节角速度 ω_t。此外，状态还会拼接参考动作中的三个未来目标帧，这些目标帧由相对于角色当前根节点坐标系的关节位置和关节线速度表示。

我们使用比例-微分（PD）控制器来生成目标关节朝向 $q' = q_b + q_r$，将基准朝向 q_b 与策略输出的残差 q_r 结合起来。虽然通常会使用下一帧参考动作 $\hat{q}_{t+1}$ 作为 q_b，但存在伪影时，这可能会误导策略。为此，我们提出采用混合式PD基准朝向，通过球面线性插值（SLERP）在当前姿势与下一帧参考姿势之间进行插值：$q_b = \text{SLERP}(q_t, \hat{q}_{t+1}, \alpha)$，其中 $\alpha \in [0, 1]$ 是策略预测的一个一维标量。

自适应终止。我们对动作模仿流程进行了多项修改，以便更好地处理不完美的参考动作。我们在世界坐标系中计算模拟角色与参考动作之间的关节位置跟踪误差，定义为 $d_p^j = ||p_t^j - \hat{p}_t^j||_2$，对应关节 j。当任意 d_p^j 超过阈值时，回合即终止。对于干净数据，固定阈值 0.5m 效果良好；但当参考动作受到干扰时，这一阈值可能过于严格且缺乏鲁棒性。为此，我们根据估计的穿透深度 ρ_t 缩放 d_p^j，从而降低受影响帧中的惩罚。具体而言，我们将位置跟踪误差放宽为：$d_p^j = \max(d_p^j - \kappa * \rho_t, 0)$，其中 $\kappa = 3.0$ 为我们实现中使用的值。因此，即使存在伪影，提前终止也会更加宽松，同时仍能生成物理上合理的动作。

奖励。我们的奖励函数鼓励精确跟踪，同时保持动作的真实性。它包含三个主要项：关节方向差异、缩放后的关节位置误差 d_p^j 以及能量正则化。缩放后的位置误差 d_p^j 还会调节奖励和终止条件，使策略能够在出现伪影时进行适应。为解决虚假接触问题，我们添加了一项用于跟踪交互动态物体运动的项。更多详情见补充材料 C。

5 多角色运动清理

多角色控制。我们将此问题建模为一个多智能体深度强化学习问题，并采用多智能体近端策略优化（MAPPO） [Yu et al. 2022]。每个角色 m 都有自己的策略 $\pi_m(s_t^m)$，动作定义与单角色情况相同。一个集中式价值函数 $V(s_t^{\text{shared}})$ 作为全局评论家。智能体分别接收个体奖励和终止信号；如果任何角色违反自适应终止阈值，则触发同步重置。我们最初尝试用单一策略控制所有角色。虽然这种方法在某些情况下有效，但在涉及严重穿透的困难场景中表现不佳。我们发现，MAPPO 得出的结果更加稳定一致。

状态与终止处理。每个角色的状态 s_t^m 包括其自身状态 s_t 以及其他角色的相对关节位置 p_t^n，这些位置在 m 的根坐标系中定义。共享状态 s_t^{shared} 由拼接所有 s_t^m 构成。为确保终止时正确进行信用分配，我们采用价值自举。如果任何智能体触发提前终止，其他智能体将获得更新后的奖励 $r_t^m = r_t^m + \lambda V(s_{t+1}^{\text{shared}})$，并一同终止。这确保了 GAE() 计算的一致性，并提高了训练稳定性。

6 结果

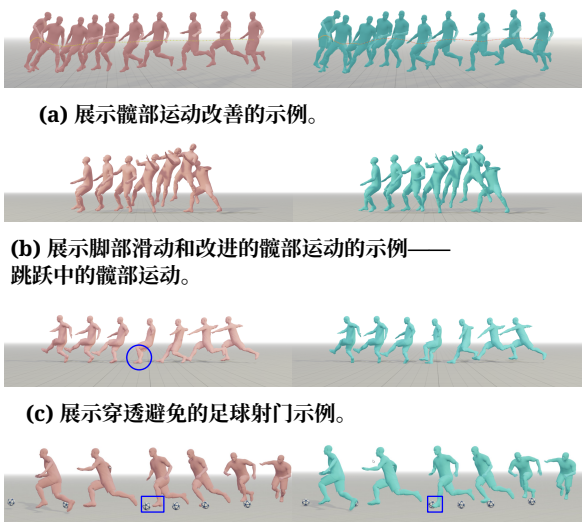
本节展示的原始动作是通过商业化的基于视觉的动作捕捉工具采集的，通常质量较高，但仍存在瑕疵。由于静态图像可能无法充分体现清理带来的改进，请参阅补充视频以便更清晰地进行比较。实现细节见补充材料 D。

6.1 运动清理

漂浮与重量感。图 6a 和 6b 对比了原始动作与精炼动作，展示了髋部运动的增强效果；图 6a 还绘制了根轨迹。这是与图 3 相同的示例，其中展示了垂直方向的根节点高度。精炼动作在两次足部着地之间呈现出更明显的正弦式根节点高度波动，增强了动作的重量感。相比之下，原始动作通常轨迹较为平坦，削弱了重力感和冲击感。定量改进结果也列于补充材料 B。

穿透。网格穿插是原始运动数据中的一个主要问题。如图 6c 所示，我们的框架解决了诸如大腿相交或小腿胫部交叉等自穿透问题，使动作更加流畅自然。图 1 进一步展示了我们的方法还修正了碰撞或拥抱等互动过程中的角色间穿透，从而实现干净且可信的接触。穿透度量的定量结果见补充材料 B。

幻影接触。我们的框架部分解决了幻影接触问题。如图 6d 所示，出现了一种缺少接触的



(a) 展示髋部运动改善的示例。

(b) 展示脚部滑动和改进的髋部运动的示例——跳跃中的髋部运动。

(c) 展示穿透避免的足球射门示例。

(d) 缺少脚部接触的足球运球示例。接触时刻以蓝框标出。

图 6：粉色角色表示原始动作，青绿色角色表示由我们的框架生成的动作。

伪影，发生在一个足球带球示例中。通过纳入足球轨迹跟踪，角色在带球过程中成功及时完成了踢球动作。然而，要将该解决方案扩展到各种人机交互问题，仍然是一项挑战。

脚部滑动。 我们的基于物理的方法通过模拟地面摩擦力来减少脚部滑动。由于静态图像中可能难以观察到这些改进，我们建议读者观看补充视频进行比较。补充材料 B 对清理前后的脚部滑动进行了定量分析。

6.2 用户引导的清理

除了自动清理之外，我们的方法还支持动画师引导的细化。在补充材料 A 的图 7 中，框架首先对踢球时膝部和脚部朝向不自然的原始动作进行了部分改进，但膝部弯曲仍不自然。随后，动画师插入一个关键帧来设定期望的姿势，并在其周围进行插值，但这一修改降低了踢球动力学的自然度。再次应用我们的框架后，动作变得更加流畅、富有动感，也更受动画师青睐。

或者，动画师也可以直接编辑特定帧。如补充材料 A 的图 8 所示，手动修正错误预测的腿部姿势后，产生了不自然的动态效果。我们的方法恢复了物理合理性，提升了稳定性、伸展效果和着陆真实感。有关论文中讨论的关键帧技术的可视化效果，请参阅补充视频。

7 讨论与结论

虽然我们的系统能够生成高保真修正结果，但目前每个动作片段都需要训练一个新的策略，这限制了其可扩展性。开发一种稳健且具备泛化能力的预训练策略，是未来工作的重要方向。此外，我们的方法假设参考动作在结构上是合理的。当输入数据受到严重破坏时，将我们的方法与数据驱动的修复或动作插帧方法相结合，可以提升整体动作质量。

总而言之，我们提出了一种基于物理的框架，用于修正视觉捕捉动作中物理上不合理的伪影。我们的方法改进了单角色和多角色序列，并且可以无缝集成到动画工作流中。我们还分析并总结了常见的伪影类型，这些伪影往往会妨碍动作达到可用于制作的标准，为未来的动作清理研究奠定了基础。

参考资料

Kang Chen、Yupan Wang、Song-Hai Zhang、Sen-Zhe Xu、Weidong Zhang 和 Shi-Min Hu。2021。Mocap-solver：一种用于光学动作捕捉数据的神经求解器。*ACM*图形学汇刊 *(TOG)* 40，4（2021），1–11。DeepMotion。2025。DeepMotion。https://www.deepmotion.com/。访问日期：2025-07-04。Elmer G Gilbert、Daniel W Johnson 和 S Sathiya Keerthi。1988。一种计算三维空间中复杂物体间距离的快速方法。*IEEE*机器人与自动化汇刊 4，2（1988），193–203。Hawkeye Innovations。2025。Hawkeye Innovations。https://www.hawkeyeinnovations. com/。访问日期：2025-07-04。Daniel Holden。2018。通过去噪稳健地求解光学动作捕捉数据。*ACM*图形学汇刊 *(TOG)* 37，4（2018），1–12。Buzhen Huang、Liang Pan、Yuan Yang、Jingyi Ju 和 Yangang Wang。2022。Neural mocon：用于物理上合理的人体动作捕捉的神经动作控制。载于*IEEE/CVF*计算机视觉与模式识别会议论文集。6417–6426。Angjoo Kanazawa、Jason Y Zhang、Panna Felsen 和 Jitendra Malik。2019。从视频中学习三维人体动态。载于*IEEE/CVF*计算机视觉与模式识别会议论文集。5614–5623。Libin Liu、KangKang Yin、Michiel Van de Panne、Tianjia Shao 和 Weiwei Xu。2010。基于采样的高接触动作控制。载于*ACM SIGGRAPH 2010* 论文集。ACM，纽约，NY，美国，1–10。Li-Ke Ma、Zeshi Yang、Xin Tong、Baining Guo 和 KangKang Yin。2021。利用时空边界学习和探索运动技能。载于计算机图形论坛，第 40 卷。Wiley Online Library，251–263。Move.ai。2025。Move.ai。https://www.move.ai/。访问日期：2025-07-04。Xue Bin Peng、Pieter Abbeel、Sergey Levine 和 Michiel Van de Panne。2018a。Deep-mimic：示例引导的深度强化学习，用于学习基于物理的角色技能。*ACM*图形学汇刊 *(TOG)* 37，4（2018），1–14。Xue Bin Peng、Angjoo Kanazawa、Jitendra Malik、Pieter Abbeel 和 Sergey Levine。2018b。SFV：从视频中强化学习物理技能。*ACM*图形学汇刊37，6，文章 178（2018 年 11 月），14 页。Yu Sun、Qian Bao、Wu Liu、Tao Mei 和 Michael J. Black。2023。TRACE：动态摄像机三维环境中的虚拟化身五维时序回归。载于*CVPR*。Tracab Technologies。2025。Tracab Technologies。https://tracab.com/products/tracab- technologies/。访问日期：2025-07-04。Gino Van Den Bergen。2001。三维游戏对象的邻近查询与穿透深度计算。载于游戏开发者大会，第 170 卷。209。Vickie Ye、Georgios Pavlakos、Jitendra Malik 和 Angjoo Kanazawa。2023。从野外视频中解耦人体与摄像机运动。载于*IEEE/CVF*计算机视觉与模式识别会议论文集。21222–21232。Chao Yu、Akash Velu、Eugene Vinitsky、Jiaxuan Gao、Yu Wang、Alexandre Bayen 和 Yi Wu。2022。近端策略优化在合作式多智能体博弈中的惊人效果。神经信息处理系统进展 35（2022），24611–24624。Ri Yu、Hwangpil Park 和 Jehee Lee。2021。利用动态摄像机运动的单目视频重建人体动态。*ACM*图形学汇刊40，6，文章 208（2021）。Ye Yuan、Shih-En Wei、Tomas Simon、Kris Kitani 和 Jason Saragih。2021。Simpoe：用于三维人体姿态估计的模拟角色控制。载于*IEEE/CVF*计算机视觉与模式识别会议论文集。7159–7169。Haotian Zhang、Ye Yuan、Viktor Makoviychuk、Yunrong Guo、Sanja Fidler、Xue Bin Peng 和 Kayvon Fatahalian。2023b。从转播视频中学习物理模拟网球技能。*ACM*图形学汇刊42，4，文章 95（2023 年 7 月），14 页。数字对象标识符：10.1145/3592408。Yunbo Zhang、Deepak Gopinath、Yuting Ye、Jessica Hodgins、Greg Turk 和 Jungdam Won。2023a。复杂多角色交互的模拟与动作重定向。载于*ACM SIGGRAPH 2023* 会议论文集。1–11。Victor Brian Zordan 和 Nicholas C Van Der Horst。2003。利用物理模型将光学动作捕捉数据映射到骨骼动作。载于*2003* 年 *ACM SIGGRAPH/Eurographics* 计算机动画研讨会论文集。Citeseer，245–250。