

1 文献基本信息

作者: Lewis Bridgeman, Marco Volino, Jean-Yves Guillemaut; 文献题目: 多人 3D 运动姿态估计与追踪; 刊物: 计算机视觉与模式识别; 刊号: ISSN(1063-6919); 时间: 2019; 格式引文: Bridgeman, Lewis et al. "Multi-Person 3D Pose Estimation and Tracking in Sports." CVPR Workshops (2019).

2 文献讨论的问题及其主要观点

写作目的 从视频估计人体 3D 姿势是一个经过充分研究的问题。但是, 针对运动领域的现有方法很少, 因为运动数据集对于计算机视觉算法尤其具有挑战性, 具体有挑战性的地方在于: 运动员快速运动并且他们之间存在相互接触; 运动员的外观类似; 运动员之间画面相互遮挡; 移动的并且低分辨率的画面且校准不佳。即便存在这样的困难, 估计运动员在运动中的 3D 姿势的潜在应用是广泛的。这包括性能分析, 运动捕捉以及在广播和沉浸式媒体中的新颖应用。基于这样的现状, 本文提出了一种适用于运动基于多视点视频的多人三维姿态信息跟踪方法。

研究方法 本文提出了一种贪婪算法, 以在多视图找到 2D 姿势之间的对应关系, 并利用它们来生成 3D 骨架。通过保持 2D 姿势的连通性, 这种方法可显著提高速度。并且通过分割姿势, 融合姿势和交换关节来纠正与多人姿势检测器相关的错误。最后, 本文引入一个算法来跟踪生成的 3D 骨架, 从而跟踪整个序列中的 2D 对应物。具体步骤为将多人和摄像机校准的多视点视频作为输入。使多视点视频通过姿势检测器, 姿势检测器每帧提供未排序的 2D 姿势估计。将三个连续过程应用于数据: 第一步纠正姿势检测器输出中的某些错误; 第二步纠正错误。第二步为每个 2D 姿势添加标签, 以确保视图之间的一致性; 最后, 标记的 2D 姿势将每个帧用于生成一系列跟踪的 3D 骨架。

主要结论 在本文中, 这种从多视点运动视频计算跟踪的 3D 骨骼的新方法可补偿姿势检测中的错误, 并可以识别不同视点中 2D 姿势之间的对应关系。该算法能够在 Campus 和 Shelf 数据集上实时运行, 也被证明可以有效地跟踪球员在拥挤的足球场上的丢失和嘈杂的检测结果。它不需要对演员的外表进行建模, 并且可以通过不良的相机校准和错误的姿势检测来实现良好的性能。

3 文献阅读后的分析

在卷积神经网络 (CNN) 方法被采用之前, 图像结构将姿势建模为连接的零件的集合, 并且先验约束了每个零件的相对位置或角度。零件以能量最小化的方式与图像数据对齐。基于 CNN 的姿势估计器可显著提高准确性, 并为更困难的姿势估计任务 (例如多人 2D) 提供基础姿势估计。通过融合关节置信度图和定义关节之间关系的学习向量场, 使用图形结构将 3D 骨骼拟合到 2D 关节以及使用卷积结构, 进行了单眼 3D 姿势估计。基于模型的早期跟踪方法使用多视图轮廓和颜色信息在工作室环境中最多跟踪两个人。为了捕获团队运动, 需要多人无标记运动捕捉方法。从多个视图获取 2D 姿态检测, 并使用体积投票找到所有 3D 关节位置。然而, 体积投票对不良的标定和错误的姿势检测很敏感, 这两者在体育数据集中都很常见。本文采用贪婪搜索来查找不同摄像机视图中 2D 姿势之间的对应关系, 并且能够以视频速率实现此目标。它仅依赖于几何术语, 而不是外观模型。补偿多人 2D 姿势估计中出现的一些典型错误。2D 姿势关联可用于生成 3D 骨架, 即使在具有严重遮挡和错误姿势检测的序列中也可进行时间跟踪。