

摘要

三维人体姿态估计是指从图片或视频中恢复人体关节点三维空间位置的任务。它在实际生活中有着广泛的应用。它促进了更加灵活方便的人机交互，同时也是动画渲染和电影制作行业的重要支撑。广阔的应用前景也使得三维人体姿态估计一直吸引着众多计算机视觉研究者的目光。然而将三维人体姿态估计技术从科学研究落地到实践应用中却不是一件容易的事情，复杂多变的实际场景对姿态估计算法的鲁棒性提出了严峻的考验。如何设计一个面对纷杂干扰仍可稳定运行的三维人体姿态估计系统是使其走出象牙塔，发挥实际功效的关键。本文将针对如何设计高鲁棒性的三维人体姿态估计算法进行研究，分别围绕其中的四个方面展开讨论：

(1) 如何在视频中鲁棒地定位和跟踪目标行人。视频中的目标定位和跟踪是三维人体姿态估计的第一步。针对视频中背景物体和跟踪目标外观相似易产生混淆的情况，本文提出了一种基于位置敏感的外观嵌入特征的目标跟踪方法，通过将位置信息融合进外观特征中，来精确分辨外观相似但位置不同的目标和背景物体。另外，针对跟踪目标外观可能发生剧烈变化的问题，本文提出了一种基于伪标签的模型调整方案，帮助模型更好地适应目标外观的变化，避免传统在线模型调整方案带来的误差累积问题。

(2) 如何提高三维人体姿态估计模型的泛化性。由于三维人体姿态数据难以标注，多由室内动作捕捉系统采集得到，成本高昂且步骤繁琐，目前公开的三维姿态数据集动作类型有限。现有模型在此类数据上的训练极易过拟合，难以泛化至动作类型更加丰富的实际场景中。针对上述问题，本文提出了一种姿态数据的图结构表示方法，并设计了一种新颖的局部连接网络来捕捉姿态数据的局部结构特征。该模型在具有强大表征能力的同时，又具有较高的泛化性。在多个数据集上的实验验证了该模型的有效性。

(3) 如何提高三维人体姿态估计结果的稳定性。现有三维人体姿态估计方法多通过神经网络直接回归人体关节点的三维坐标。输出结果不受任何约束，因此稳定性较差，易产生不符合人体解剖学比例的姿态估计结果。针对此问题，本文提出了一种基于上下文信息建模的方法。该方法通过注意力机制将人体肢体长度的先验知识编码进网络结构中，在保证端到端训练、利用神经网络强大数据拟合能力的同时，引导其产生符合人体解剖学比例的合理估计结果。多个数据集上的实验表明，该方法有效地增强了姿态估计结果的稳定性。

(4) 如何提高三维人体姿态估计系统的抗遮挡性能。之前的工作多关注于遮挡发生时如何增强算法对遮挡的抵抗能力。与之不同，本文探索了一种主动式的遮挡避免策略。通过将相机挂载在无人机上，主动地在三维空间中控制相机的移动来探索适合重

建的最佳无遮挡视角。本文将相机的控制问题建模为马尔可夫决策过程，借助深度强化学习求解最优控制策略。同时，为了提升策略的长期规划能力，本文将隐式的环境建模融入到模型的在线学习框架中，以进一步增强策略对环境动态的理解与预判。

另外，基于以上研究内容，本文搭建了一套智能体育教学辅助系统，将三维人体姿态估计技术落地于智能体育动作评估与姿态纠正的应用中。同时，本文将从应用和系统的角度探讨如何提升姿态估计技术在落地部署和实际运行中的鲁棒性。

关键词：计算机视觉，三维人体姿态估计，鲁棒性