

8-13-2020

3D Human Pose Tracking Approach Based on Double Kinect Sensors

Li Qi

1. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;;2. University of Chinese Academy of Sciences, Beijing 100049, China;;

Xiangdong Wang

3. China Institute of Sport Science, General Administration of Sport, Beijing 100061, China;

Li Hua

1. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;;2. University of Chinese Academy of Sciences, Beijing 100049, China;;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

3D Human Pose Tracking Approach Based on Double Kinect Sensors

Abstract

Abstract: 3D human pose estimation and tracking is a popular problem of machine vision. In order to improve the stability of human pose tracking, *the human body motion posture is represented by the human joint DOF (Degree of Freedom) vector, the human pose is tracked by the unscented Kalman filter method, and a human pose tracking system based on double Kinect sensors is built.* Compared with the traditional human motion capture system, the system can accurately and stably track 3D human pose under simple movements, reflect the special nature of the motion process under complex movements, and can be used in the performance evaluation of sports biomechanics.

Keywords

3D human pose tracking, Kinect sensor, human joint DOF, unscented Kalman filter, sports biomechanics

Recommended Citation

Li Qi, Wang Xiangdong, Li Hua. 3D Human Pose Tracking Approach Based on Double Kinect Sensors[J]. Journal of System Simulation, 2020, 32(8): 1446-1454.

基于双 Kinect 传感器的三维人体姿态跟踪方法

李琪^{1,2}, 王向东³, 李华^{1,2}

(1.中国科学院计算技术研究所 院智能信息处理重点实验室, 北京 100190; 2.中国科学院大学, 北京 100049;
3. 国家体育总局体育科学研究所, 北京 100061)

摘要: 三维人体姿态估计与跟踪是机器视觉中的热点问题。为了提高人体姿态跟踪的稳定性, 以人体关节点自由度向量的方法表示人体运动姿态, 使用无迹卡尔曼滤波方法对人体姿态进行跟踪, 并搭建了一套基于双 Kinect 传感器的人体姿态跟踪系统。对比于传统的人体运动捕捉系统, 该系统可实现在简单动作下准确稳定的三维人体姿态跟踪, 在复杂动作下反映运动过程的特殊性质, 并用于运动生物力学的性能评估。

关键词: 三维人体姿态跟踪; Kinect 传感器; 关节点自由度; 无迹卡尔曼滤波; 运动生物力学

中图分类号: TP391 文献标识码: A 文章编号: 1004-731X (2020) 08-1446-09

DOI: 10.16182/j.issn1004731x.joss.19-0017

3D Human Pose Tracking Approach Based on Double Kinect Sensors

Li Qi^{1,2}, Wang Xiangdong³, Li Hua^{1,2}

(1. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China;
3. China Institute of Sport Science, General Administration of Sport, Beijing 100061, China)

Abstract: 3D human pose estimation and tracking is a popular problem of machine vision. In order to improve the stability of human pose tracking, the human body motion posture is represented by the human joint DOF (Degree of Freedom) vector, the human pose is tracked by the unscented Kalman filter method, and a human pose tracking system based on double Kinect sensors is built. Compared with the traditional human motion capture system, the system can accurately and stably track 3D human pose under simple movements, reflect the special nature of the motion process under complex movements, and can be used in the performance evaluation of sports biomechanics.

Keywords: 3D human pose tracking; Kinect sensor; human joint DOF; unscented Kalman filter; sports biomechanics

引言

三维空间中的人体姿态跟踪指的是估计人体关节点的三维坐标, 并保持对人体运动姿态的跟

踪。三维人体姿态跟踪是机器视觉中的热点问题, 其后续应用包括动作识别、人机交互与视觉理解。传统的三维人体姿态跟踪(或称为运动捕捉)方法是在人体关节点表面贴上具有荧光性质的标记点, 使用红外光对这些标记点进行跟踪, 进而获取人体运动状态。虽然这种方法对于人体关节点定位精确, 但是它具有设备价格昂贵和搭建准备复杂的缺点, 在人体表面贴标记点的约束也大大限制了这一方法的适用环境。为了对一名受试者的人体姿态进



收稿日期: 2019-01-09 修回日期: 2019-03-12;
基金项目: 国家重点研发计划(2017YFB1002703), 国家重点基础研究规划(2015CB554507), 国家自然科学基金(61379082), 国家体育总局体育科学研究所合作项目(基本 17-01, 基本 17-28, 基本 18-19);
作者简介: 李琪(1992-), 男, 安徽蚌埠, 博士生, 研究方向为计算机图形学与人体动作识别。

<http://www.china-simulation.com>

• 1446 •

行跟踪分析, 往往需要 20~30 min 的准备时间。本文提出了一种基于双 Kinect 传感器的三维人体姿态跟踪方法, 并搭建了一套完整的跟踪系统。相对于传统的运动捕捉方法, 基于双 Kinect 的跟踪系统仅需 1 min 左右的标定准备即可使用, 可用于实验室环境下快速的三维人体姿态跟踪。

1 相关工作与简介

基于视频图像或其他传感器的三维人体姿态估计与跟踪有着长期的研究历史。Mousland 等^[1]总结了早期的基于视频图像的三维人体姿态跟踪的相关研究, 其中基于模型的匹配方法较为常见。这一类方法将显式的三维人体模型与二维图像特征进行匹配以确定人体关节的三维坐标, 但是这类方法有着计算量大且容易陷入局部最优而跟踪失败的问题, 在现在的研究中已经较为少见。

研究者们尝试使用多种传感器以收集可以表征人体姿态的信息, 其中包括惯性测量单元, 激光测距传感器和 RGBD(RGB+Depth map)传感器。以微软 Kinect 为主要代表的 RGBD 传感器不仅包括常用的彩色摄像头, 还包括一个使用红外光进行测距的深度摄像头。Shotton 等^[2]基于 Kinect 获取的深度图像, 使用随机森林的方法回归求取人体关节的三维坐标。Wei 等^[3]将深度摄像头获取的三维点云与人体模型进行拟合, 不断优化以获取准确的人体关节三维坐标, 但其人体模型较为简单, 躯干和四肢以简单的圆柱体进行表示。Shuai 等^[4]对这一方法做了进一步的拓展, 利用多台 Kinect 获取的深度点云的融合数据, 将人体表面模型以球面谐波函数的方式获取与拟合, 进一步提高了关节位置的精度。现阶段, Kinect 传感器已自带 Shotton 等的算法实现, 但是其精度不够高, 仅适用于体感游戏中的人体姿态跟踪。Guess 等^[5]对特定运动中 Kinect 2.0 与传统人体运动捕捉设备获取的人体关节角度进行了对比分析, 认为单台 Kinect 2.0 所获取的人体关节位置精度是有限的。Moon 等^[6]利用多台 Kinect 2.0 获取的人体关节位置, 使用基

于卡尔曼滤波的数据融合方法, 获取更为精确与稳定的人体关节轨迹。然而, 他们的算法将每个人体关节单独考虑并绘制它们的轨迹, 忽略了关节之间的位置关系。不仅如此, 为了进一步提高算法精度, Moon 等增加了 Kinect 传感器的数量。当使用微软官方的驱动时, 每台计算机只能连接一台 Kinect 2.0 传感器, 并且 Kinect 2.0 传感器没有同步线接口。因此当增加 Kinect 传感器数量时, 整个传感器系统的成本和标定同步等配置操作都会显著增加。Joo 等^[7]利用普通与高速摄像机以及 Kinect 传感器搭建了一个封闭环境, 可精确估计并跟踪多人三维人体姿态, 但整个系统造价昂贵, 需要大量的摄像机与传感器设备从 480 个视角捕捉人体运动。

随着深度学习方法的广泛使用, 研究者尝试使用神经网络解决三维人体姿态估计的问题。然而, 由于三维人体姿态数据库^[8]仅包含室内场景, 动作种类也有所局限, 在这些数据库上训练的模型估计三维人体姿态的精确性与泛化性都受到了较大的约束^[9]。不仅如此, 基于深度学习的三维人体姿态估计算法在训练与推测阶段对于机器算力都有较高的要求, 现阶段还难以应用于人体姿态跟踪中。

本文算法对 Moon 等的工作进行了改进, 仅利用双 Kinect 2.0 传感器获取的人体关节坐标进行数据融合, 就可以获得准确而稳定的人体关节位置轨迹。本文算法基于各关节的自由度^[10-11]定义了一个表示人体运动状态的状态向量, 利用无迹卡尔曼滤波器^[12]进行状态向量的预测与更新。基于自由度的表示方法利用根节点的坐标和关节角度表示人体运动状态, 对相邻关节之间的距离即骨节长度基于预采集的数据进行估计, 这种表示方法既保证了关节位置随观测值修正的灵活性, 又保证了关节之间位置关系的稳定性。相比于 Moon 等算法中使用的卡尔曼滤波, 无迹卡尔曼滤波器可以处理非线性的状态转移与变换。进一步地, 本文实现了一个基于双 Kinect 传感器的三维人体姿态跟踪系统, 两台 Kinect 2.0 连接在同一台计算机上,

减少了传感器之间标定与同步的操作。

2 基于无迹卡尔曼滤波的人体姿态跟踪算法研究

2.1 整体流程

本文方法的整体流程如下：首先对双 Kinect 传感器进行标定与同步，进而对 Kinect 采集到的人体关节数据点进行空间转换、时间同步、数据融合与平滑滤波，最后计算输出运动过程中对应关节角度随时间的变化情况。

2.2 双 Kinect 传感器的标定与同步

在三维人体姿态跟踪之前，需要对 Kinect 传感器进行标定与同步。对每台 Kinect 传感器的深度摄像头进行平板标定后，基于式(1)计算 2 个摄像机坐标系之间的关系：

$$\begin{aligned} P_l &= R_l P + T_l \\ P_r &= R_r P + T_r \\ P_l &= R^T (P_r - T_r) \\ R &= R_r R_l^T \\ T &= T_r - R_r R_l^T T_l \end{aligned} \quad (1)$$

式(1)表示给定世界坐标系下 P 点，将其转换至左右深度摄像机坐标系中的转换关系。根据左右深度摄像机的外参数 R_l , T_l 与 R_r , T_r ，可以计算左右摄像机坐标系之间的变换关系 R 和 T 。Kinect 传感器并没有同步线的设置，因此不能直接通过硬件方法对双 Kinect 进行同步。本文采用的方法是当采集到特定帧的人体关节数据时，将对应的计算机系统时间记录下来，然后基于系统时间进行线性插值。逐个比较双 Kinect 采集关节的对应时刻，对于当前帧数相同的 2 个对应时刻，取其中较小值为 t_1 ，则另一个时刻值为 t_2 。 t_2 所对应传感器的上一次采集时刻设为 t_{prev} ，则可将 t_2 所对应的关节三维坐标进行线性插值，以求取它在 t_1 时刻的具体位置。其公式为：

$$\begin{aligned} pt_{t_1} &= pt_{t_{prev}} + \\ & (pt_{t_2} - pt_{t_{prev}}) \times (t_1 - t_{prev}) / (t_2 - t_{prev}) \end{aligned} \quad (2)$$

2.3 人体姿态表示方法

Kinect 传感器给出的关节三维坐标仅能表示关节的空间位置，不能有效地表现关节之间的位置关系。本文采用以髋为根节点的树结构表示人体姿态，如图 1 所示。每个关节(除根节点)的自由度由其可活动的欧拉角数目所定义，根节点的自由度还包括其空间位置的自由度，由图 1 定义的人体姿态的自由度为 24。本文把每个自由度的变化率也加入状态向量中。每个时刻的状态向量 x_t 的前 24 维用关节角自由度表示，后 24 维由与关节角自由度相对应的变化率表示，因此 $x_t \in \mathbb{R}^{48}$ ，且 x_t 唯一确定了时刻 t 人体的运动状态。

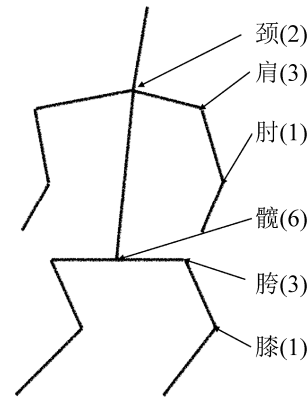


图 1 人体骨架模型及对应关节的自由度

Fig. 1 Human skeleton model and DOF of corresponding joints

在该骨架模型中，各骨节长度是需要预先设置的参数，其估计方法是对双 Kinect 预采集的一系列的人体姿态样本求取各骨节长度，在除去明显的异常值后求取样本均值，并作为该人体姿态的参数存储下来。

2.4 基于无迹卡尔曼滤波器的融合平滑算法

2.4.1 无迹卡尔曼滤波

由于基于关节自由度的姿态表示方法与 Kinect 输出的关节三维坐标存在着非线性变换，传统的卡尔曼滤波不能有效地对这种非线性变换进行建模。在卡尔曼滤波算法及其变种算法中，状态变量与测量变量均假设为满足高斯分布的随

机变量,而非线性函数接收自变量随机变量的均值为输入时,其输出并非因变量随机变量的均值,而会产生一定的偏移,这被称为均值漂移现象。为了解决这一问题,无迹卡尔曼滤波是对卡尔曼滤波在非线性的推广,基本假设是卡尔曼滤波中的状态转移矩阵 F 和测量转移矩阵 H 均变为非线性函数 $f(\mathbf{x})$ 和 $h(\mathbf{x})$,基本方法是通过采样的方法求解估计随机变量经非线性函数变换后的均值与协方差。无迹卡尔曼滤波在每一步的状态更新中有 2 个阶段:预测阶段与更新阶段。

预测阶段中,基于 t 时刻的状态向量 \mathbf{x}_t 和协方差矩阵 \mathbf{P}_t ,使用采样算法^[13]获取 $2n+1$ 个采样点 $\{\mathbf{X}_i, i=0, 1, \dots, 2n\}$ 及其对应的权重 $\{w_i^m, w_i^c, i=0, 1, \dots, 2n\}$,其中 n 表示状态向量 \mathbf{x} 的维数。将采样点 $\{\mathbf{X}_i\}$ 代入 $f(\mathbf{x})$ 中可求解对应的函数值 $\{\mathbf{Y}_i\}$,于是 $t+1$ 时刻的预测的状态向量与协方差矩阵表示为:

$$\hat{\mathbf{x}}_{t+1} = \sum_{i=0}^{2n} w_i^m \mathbf{Y}_i \quad (3)$$

$$\hat{\mathbf{P}}_{t+1} = \sum_{i=0}^{2n} w_i^c (\mathbf{Y}_i - \hat{\mathbf{x}}_{t+1})(\mathbf{Y}_i - \hat{\mathbf{x}}_{t+1})^T + \mathbf{Q} \quad (4)$$

更新阶段中,需要首先计算样本点 $\{\mathbf{Y}_i\}$ 经过非线性函数 $h(\mathbf{x})$ 后得到的 $\{\mathbf{Z}_i\}$ 的样本均值与方差:

$$\mathbf{u}_z = \sum_{i=0}^{2n} w_i^m \mathbf{Z}_i \quad (5)$$

$$\mathbf{P}_z = \sum_{i=0}^{2n} w_i^c (\mathbf{Z}_i - \mathbf{u}_z)(\mathbf{Z}_i - \mathbf{u}_z)^T + \mathbf{R} \quad (6)$$

进一步地,计算残差向量与卡尔曼增益:

$$\mathbf{y} = \mathbf{z} - \mathbf{u}_z \quad (7)$$

$$\mathbf{K} = \left[\sum_{i=0}^{2n} w_i^c (\mathbf{Y}_i - \hat{\mathbf{x}}_{t+1})(\mathbf{Z}_i - \mathbf{u}_z)^T \right] \mathbf{P}_z^{-1} \quad (8)$$

更新的 $t+1$ 时刻的状态向量 \mathbf{x}_{t+1} 和协方差矩阵 \mathbf{P}_{t+1} 按式(9)与式(10)计算:

$$\mathbf{x}_{t+1} = \hat{\mathbf{x}}_{t+1} + \mathbf{K}\mathbf{y} \quad (9)$$

$$\mathbf{P}_{t+1} = \hat{\mathbf{P}}_{t+1} + \mathbf{K}\mathbf{P}_z\mathbf{K}^T \quad (10)$$

需要注意的是,公式(4)和(6)的转移噪声协方差矩阵 \mathbf{Q} 和测量噪声协方差矩阵 \mathbf{R} 的处理方式是不一样的, \mathbf{Q} 设置为一个定值,如单位矩阵 \mathbf{I} 。本文算法通过对测量噪声协方差矩阵 \mathbf{R} 的估计实现

双 Kinect 测量数据的加权平均。非线性函数 $f(\mathbf{x})$ 与 $h(\mathbf{x})$ 分别表示状态向量的预测转移函数与状态向量转化为测量向量的函数。在本文方法中, $f(\mathbf{x})$ 是一个线性函数,其矩阵形式可定义为:

$$\mathbf{F} = \begin{bmatrix} \mathbf{I} & \Delta t \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad (11)$$

式中: \mathbf{I} 为维数为 24 的单位阵; $\mathbf{0}$ 为维数为 24 的零矩阵; Δt 为预测转移的时间差。该函数表示以自由度表示的状态向量 \mathbf{x}_t 中,在预测阶段各自由度根据变化率与时间差发生变化,而变化率保持不变。 $h(\mathbf{x})$ 是根据以自由度表示的状态向量 \mathbf{x}_t 求解各关节的三维坐标的非线性函数,以轴角表示关节的自由度,通过正运动学方法求解各关节的三维坐标,详情可参见文献[10-11]。

2.4.2 测量噪声向量与协方差矩阵的估计

时刻 t 的双 Kinect 的测量向量表示为 $\{z_t^1, z_t^2\}$, 这里的 z 由获取的关节的三维坐标级联而成,如当关节数目为 15 个时,该测量向量的维数为 45。基于测量向量 $\{z_t^i, i=1, 2\}$ 遵循分布于样本均值周围的高斯分布的假设,当测量值距样本均值越近时,其测量噪声越小,当测量值距离样本均值越远时,其测量噪声越大,可估计测量噪声向量 \mathbf{v}_t 计算为测量向量取值为 $\{z_t^i, i=1, 2\}$ 时的概率密度函数的倒数:

$$\mathbf{v}_t = \frac{1}{p(z_t^i = z_t^i | \mathbf{u}_z)}, \text{ when} \\ p(z_t^i = z_t^i | \mathbf{u}_z) \sim \mathbf{N}(\mathbf{u}_z, \sum_{i=0}^{2n} w_i^c (\mathbf{Z}_i - \mathbf{u}_z)(\mathbf{Z}_i - \mathbf{u}_z)^T) \quad (12)$$

进一步地,基于 \mathbf{v}_t 可对测量噪声协方差矩阵 \mathbf{R} 进行估计,假设 \mathbf{v}_t 的各分量相互独立,可估计 \mathbf{R} 为对角阵,其对角元素计算为 \mathbf{v}_t 对应分量的平方:

$$\mathbf{R} = \begin{cases} \mathbf{R}_{ii} = E\{\mathbf{v}\mathbf{v}^T\} \\ \mathbf{R}_{ij} = 0 \end{cases} \quad (13)$$

2.4.3 加权融合与平滑

这里对双 Kinect 的测量向量基于噪声的大小进行加权。当测量噪声越小,表示测量值距离样本均值越近,其权重应该越高;当测量噪声越大,表

示测量值距离样本均值越远, 其权重应该越低。因此, 以测量噪声协方差矩阵 \mathbf{R} 的逆矩阵为权重因子, 对双 Kinect 的测量向量进行加权平均:

$$\mathbf{R}_t = ((\mathbf{R}_t^1)^{-1} + (\mathbf{R}_t^2)^{-1})^{-1} \quad (14)$$

$$\mathbf{z}_t = \mathbf{R}_t((\mathbf{R}_t^1)^{-1}\mathbf{z}_t^1 + (\mathbf{R}_t^2)^{-1}\mathbf{z}_t^2) \quad (15)$$

将式(14)代入式(6), 将式(15)代入式(7), 即可基于双 Kinect 传感器采集的人体关节数据融合平滑, 跟踪三维人体运动姿态。

3 实验结果与分析

3.1 实验设置

为评估本文方法对三维人体姿态的跟踪效果, 本文使用开发的双 Kinect 三维人体姿态跟踪系统与传统的三维人体运动捕捉系统 Vicon 同时采集一段运动数据并进行对比评估。双 Kinect 传感器的布置如下: 一台 Kinect 正面面向受试者, 另一台有一定的移动与倾斜, 由于 Kinect 中内嵌的 Shotton 等的算法^[2]以正面姿态为主要训练数据, 因此其正面姿态估计精确度相比侧面姿态估计精确度要高, 所以本文没有设计放在人体侧面的 Kinect。双 Kinect 的好处在于当其中一台 Kinect 由于自遮挡等问题对于关节位置估计不稳定时, 另一台可以提供相对精确的关节位置, 为后续的融合平滑算法提供相对精确的数据。受试者的活动区域约为边长 2.5 m 的矩形区域, 活动区域中心距 Kinect 的距离约为 2 m, 离地面的高度约为 1.3 m。系统采集的人体运动姿态如图 2 所示。

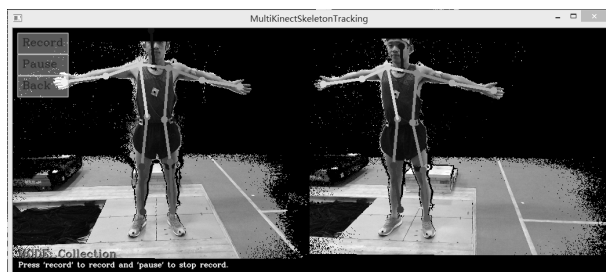


图 2 双 Kinect 三维人体姿态跟踪系统图示

Fig. 2 3D human pose tracking system based on double Kinect sensors

实验中采用的 Kinect 传感器是微软推出的第二代 Kinect。微软官方驱动仅支持一台计算机连接一台 Kinect 2.0, 不利于快速低成本地搭建多 Kinect 系统。因此, 本文开发的系统中使用由多伦多大学开发并维护的驱动及相关 API: libfreenect2, OPENNI2 和 NITE2, 并将双 Kinect 与同一台计算机连接, 降低了传感器之间标定与同步的要求。这套驱动与 API 的缺点在于其提供的关节数目相比官方驱动较少, 仅有 15 个, 但主要人体关节如肩肘腕、跨膝踝、颈部和头部都包含在内, 因此对于本文系统的构建影响较小。在实验中, 本文发现当连接 Kinect 数目多于 2 台时, 采集数据会出现明显的丢包现象, 大幅降低人体姿态跟踪精度, 因此最终选用双 Kinect 构建三维人体姿态跟踪系统, 整个系统的帧率在 25 帧左右。

需要注意的是, 由于人体姿态呈现高度的复杂性, Moon 等^[6]使用 5 台 Kinect 从多个视角采集并不能显著提高跟踪精度。由 Joo 等^[7]的实验可知, 当视角增加到 100 个以上时, 多视角系统的精度才会显著增加。因此, 考虑到成本与便携性, 本文系统使用双 Kinect 传感器是可行的。然而, 本文算法可以进一步推广到视角更多的情况, 用于多传感器的融合平滑与滤波, 以估计并跟踪人体运动, 且时间复杂度随着传感器数量线性增加。

实验采集 2 种不同的动作: (1) 侧抬腿后跳起动作, 需要受试者首先站在具有一定高度的木板箱上, 然后向侧方向抬起右腿。收回右腿后, 从木板箱上跳下, 然后完成原地起跳动作; (2) 侧抬腿后原地踏步动作, 需要受试者站在活动区域中心, 向侧方向抬起右腿, 然后进行原地踏步动作。以关节轨迹的平滑程度评估跟踪方法的稳定性, 以双 Kinect 系统与传统人体运动捕捉系统 Vicon 获取的人体关节角度对比评估跟踪方法的准确性。由于 2 种动作均以下肢动作为主, 因此评估中使用的关节角度为膝关节的外展角、弯曲角与膝关节的弯曲角。双 Kinect 系统与 Vicon 系统没有进行标定与同步, 由于侧抬腿动作中右膝关节的外展角有显著变

化,使 2 种系统获取的侧抬腿阶段的右膝关节外展角手动对齐,可用于同步 2 种系统获取的关节角度。

类似于文献[5],为评估双 Kinect 系统与 Vicon 系统获取的关节角度的一致性与差别,本文选用 2 种指标进行定量评估,这 2 种指标分别为皮尔逊相关系数(Pearson's correlation coefficient)与均方误差(Mean square error)。皮尔逊相关系数表示 2 个关节角序列的相关性,取值在-1 和 1 之间,越高说明 2 个序列的正相关性越强。均方误差表示 2 个关节角序列的平均差异,单位是度,越小说明 2 个序列越一致。本文绘制不同方法获取的关节角度随时间的变化情况,以定性观察哪种方法的关节角度更贴近于 Vicon 系统。

3.2 实验分析

以一名受试者的右膝关节的三维坐标在侧抬腿后跳起动作中随时间的变化关系举例,如图 3 所示。可以看出,单台 Kinect 获取的人体关节点坐标易出现异常值,这往往是自遮挡导致的。Moon 等提出的卡尔曼滤波算法能在一定程度上消除异常值的影响,但本文基于无迹卡尔曼滤波的方法在平滑性上进一步提高,主要由于显式地定义了关节点之间的位置关系,从而使得跟踪的人体运动姿态更为稳定。

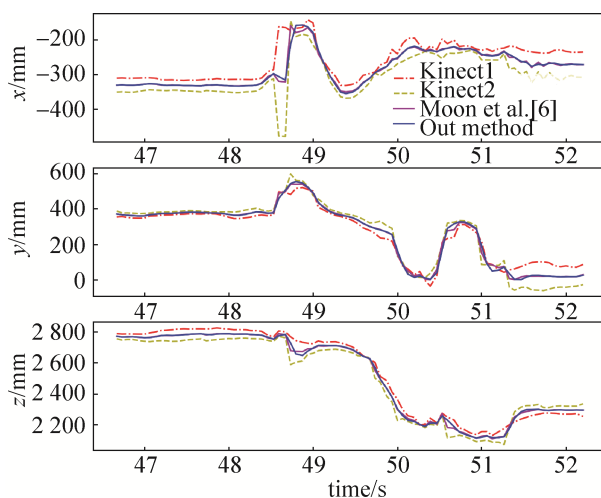


图 3 右膝关节点 x , y , z 坐标随时间的变化关系
Fig. 3 x , y and z coordinates change of right knee joint

本文对比了该受试者在 2 套动作中双 Kinect 系统和 Vicon 系统获取的关节角度随时间的变化情况。如表 1~4 所示,分别为 2 种不同动作下各方法所获取的关节角序列与 Vicon 系统获取的关节角度的相关系数与均方误差,加粗表示各方法中的最优值。从表 1 和表 2 可以看出,在较简单的动作——侧抬腿后跳起动作中,本文方法略优于 Moon 等的方法,且多数情况下比原始 Kinect 采集的数据要好。从表 3 和表 4 可以看出,在较复杂的动作——侧抬腿后原地踏步动作中,原始 Kinect 采集的数据相对表现较差,本文方法仍能较好地提升与实际关节角序列的相关性,如右膝关节外展角和左膝关节弯曲角。多数情况下,对比于 Moon 等的方法,本文方法在相关性提升和误差降低方面的表现更好,这说明在复杂动作情况下,本文方法更能体现出运动过程的特性,如周期性等。在简单动作下,本文方法的相关系数在 0.8 左右,均方误差约在 10° 左右,仍能较好地估计每一时刻的实际关节角度。在复杂动作下,本文方法的相关系数波动较大,均方误差也明显提高,仅可用于估计运动过程中关节角度的变化情况。

表 1 侧抬腿后跳起动作各方法关节角度的相关系数
Tab. 1 Correlation coefficient of joint angles of each method in sliding leg lift, jumping up and down

关节角	Kinect 1	Kinect 2	Moon et al. ^[6]	Ours
右膝关节弯曲角	0.811	0.818	0.868	0.890
左膝关节弯曲角	0.828	0.814	0.850	0.871
右膝关节外展角	0.829	0.904	0.911	0.918
左膝关节外展角	0.721	0.590	0.731	0.719
右膝关节弯曲角	0.644	0.541	0.707	0.738
左膝关节弯曲角	0.833	0.896	0.893	0.874

表 2 侧抬腿后跳起动作各方法关节角度的均方误差
Tab. 2 Mean square error of joint angles of each method in sliding leg lift, jumping up and down

关节角	Kinect1	Kinect2	Moon et al. ^[6]	Ours
右膝关节弯曲角	13.1	13.9	13.0	11.6
左膝关节弯曲角	21.8	22.6	22.1	21.4
右膝关节外展角	7.6	9.8	7.4	7.5
左膝关节外展角	7.1	10.7	8.3	8.4
右膝关节弯曲角	18.0	21.3	17.3	16.9
左膝关节弯曲角	14.4	15.3	14.4	14.4

表 3 侧抬腿后原地踏步动作各方法关节角度的相关系数
Tab. 3 Correlation coefficient of joint angles of each method in sliding leg lift and marketing time

关节角	Kinect1	Kinect2	Moon et al. ^[6]	Ours
右膝关节弯曲角	0.369	0.749	0.753	0.719
左膝关节弯曲角	0.548	0.691	0.742	0.761
右膝关节外展角	0.425	0.552	0.519	0.655
左膝关节外展角	0.714	0.284	0.643	0.650
右膝关节弯曲角	0.334	0.157	0.409	0.400
左膝关节弯曲角	0.126	0.250	0.300	0.479

表 4 侧抬腿后原地踏步动作各方法关节角度的均方误差
Tab. 4 Mean square error of joint angles of each method in sliding leg lift and marketing time

关节角	Kinect1	Kinect2	Moon et al. ^[6]	Ours
右膝关节弯曲角	40.5	35.3	28.7	29.7
左膝关节弯曲角	18.0	20.4	17.8	16.8
右膝关节外展角	14.3	12.6	10.6	9.2
左膝关节外展角	18.1	21.9	19.6	19.5
右膝关节弯曲角	67.3	83.7	65.0	63.3
左膝关节弯曲角	39.8	50.5	32.6	30.0

如图 4~5 所示, 在侧抬腿后跳起动作中, 由于动作较为简单, 双 Kinect 系统能较精确地跟踪人体运动, 反映运动过程中关节角度的变化情况, 获取的关节角度与 Vicon 系统获取的关节角度偏差较小。在侧抬腿后原地踏步动作中, 双 Kinect 系统获取的关节角度能反映出运动过程的周期性, 但没有 Vicon 系统获取的稳定, 关节角度与 Vicon 系统的关节角的偏差也较大。因此, 在简单动作情况下, 双 Kinect 系统已能较精确地跟踪三维人体运动状态; 在复杂动作情况下, 即使人体姿态有较多自遮挡, 双 Kinect 系统仍能反映出运动过程的周期性等特殊性质, 并用于运动训练的性能评估。

由图 4 和图 5 可以进一步得知, 单台 Kinect 获取的关节角度不稳定性较高, 且易于与 Vicon 系统获取的关节角度产生较大的偏差, 这主要是由于 Kinect 传感器易受到人体姿态自遮挡的影响。

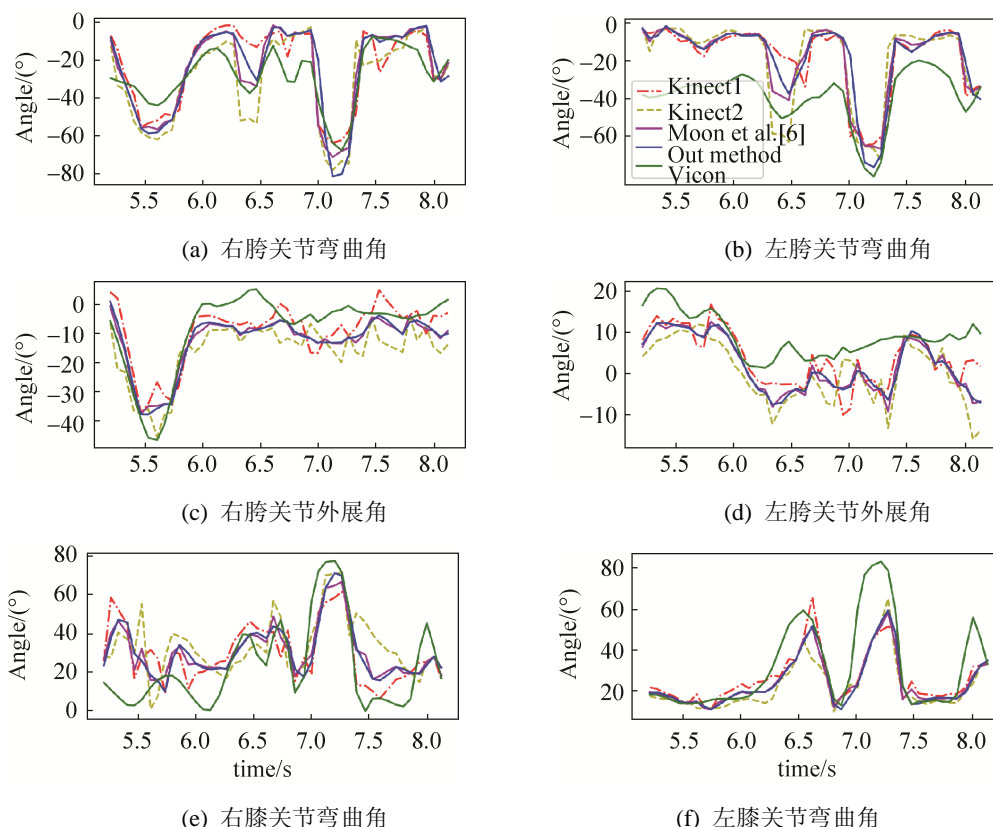


图 4 侧抬腿后跳起动作各关节角的变化情况
Fig. 4 Change of joint angles in sliding leg lift, jumping up and down

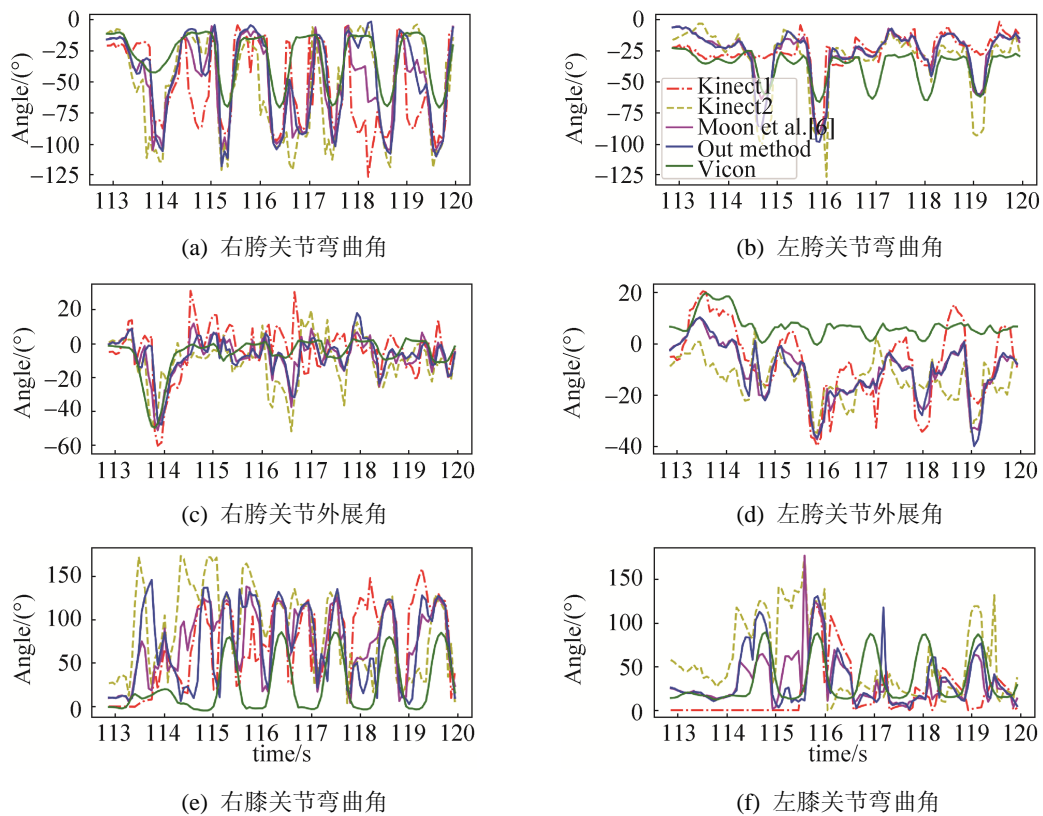
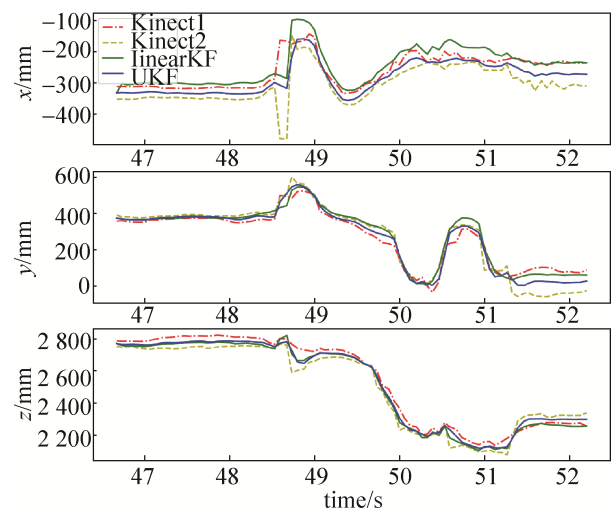


图 5 侧抬腿后原地踏步动作各关节角的变化情况

Fig. 5 Change of joint angles in sliding leg lift and marketing time

在 2 种动作中, 本文方法与 Moon 等^[6]的方法对于关节角度变化趋势的刻画是趋于一致的, 并和 Vicon 系统采集的关节角度的变化情况相似, 这说明了加权融合平滑算法的有效性。相比于 Moon 等的方法, 本文方法获取的关节角度在多数关节角的波峰位置与 Vicon 系统获取的关节角度的偏差更小一些, 这说明本文所使用的基于自由度的人体姿态表示方法比直接使用关节点三维坐标的表示方法更为准确。为了验证无迹卡尔曼滤波的必要性, 本文设置了一个实验, 将关节点三维坐标的表示非线性转换为基于关节点自由度的表示, 然后再进行传统的卡尔曼滤波过程, 其跟踪的人体姿态结果如图 6 所示, 可以看出右膝关节的 x 与 y 坐标产生了明显的偏移, 偏出了左右 Kinect 采集的关节点三维坐标的中间范围。由于本文算法是对左右 Kinect 采集结果的融合加权, 这说明在将关节点三维坐标的表示非线性转换为基于关节点自由度的表示时出

现了均值偏移的现象, 从而使得加权融合的结果偏离了正常的范围。

图 6 使用线性卡尔曼滤波的情况下右膝关节点 x , y , z 坐标随时间的变化关系Fig. 6 x , y and z coordinates change of right knee joint using linear Kalman filter

当 2 台 Kinect 估计的人体姿态均不够准确时, 双 Kinect 系统也会给出不准确的三维人体姿态。在本文的实验过程中, 出现一名受试者在落地时因平衡问题稍微后仰, 而 2 台 Kinect 均没有估计出其准确的人体姿态, 因此本文的双 Kinect 加权融合的三维人体姿态也是不准确的。这一问题可以通过增加 Kinect 传感器的数目加以缓解, 但并没有从根本上解决其问题。准确的三维人体姿态估计仍是未来研究的重点, 是高精度的人体姿态跟踪的前提。

4 结论

本文算法基于人体关节自由度表示的人体运动姿态与无迹卡尔曼滤波方法, 对于双 Kinect 传感器估计的人体姿态数据进行加权融合与平滑, 实现了简单动作中相对精确稳定的三维人体姿态跟踪方法。本文算法在复杂动作情况下仍能反映运动的特殊性质, 并用于运动训练的运动评估。本文将尝试发掘更为精确的三维人体姿态估计方法, 并提高三维人体姿态跟踪的精度。

参考文献:

- [1] Moeslund T B, Hilton A, Krüger V. A Survey of Advances in Vision-based Human Motion Capture and Analysis[J]. *Computer Vision and Image Understanding* (S1077-3142), 2006, 104(2/3): 90-126.
- [2] Shotton J, Girshick R, Fitzgibbon A, et al. Efficient Human Pose Estimation from Single Depth Images[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (S0162-8828), 2013, 35(12): 2821-2840.
- [3] Wei X, Zhang P, Chai J. Accurate Real-time Full-body Motion Capture Using a Single Depth Camera[J]. *ACM Transactions on Graphics* (S0730-0301), 2012, 31(6): 188.
- [4] Shuai L, Li C, Guo X, et al. Motion Capture with Ellipsoidal Skeleton Using Multiple Depth Cameras[J]. *IEEE Transactions on Visualization and Computer Graphics* (S1077-2626), 2017, 23(2): 1085-1098.
- [5] Guess T M, Razu S, Jahandar A, et al. Comparison of 3D Joint Angles Measured with the Kinect 2.0 Skeletal Tracker versus a Marker-based Motion Capture System[J]. *Journal of Applied Biomechanics* (S1065-8483), 2017, 33(2): 176-181.
- [6] Moon S, Park Y, Ko D W, et al. Multiple Kinect Sensor Fusion for Human Skeleton Tracking Using Kalman Filtering[J]. *International Journal of Advanced Robotic Systems* (S1729-8814), 2016, 13(2): 65.
- [7] Joo H, Simon T, Li X, et al. Panoptic Studio: A Massively Multi-view System for Social Interaction Capture[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (S0162-8828), 2019, 41(1): 190-204.
- [8] Ionescu C, Papava D, Olaru V, et al. Human3.6m: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (S0162-8828), 2014, 36(7): 1325-1339.
- [9] Zhou X, Huang Q, Sun X, et al. Towards 3D Human Pose Estimation in the Wild: A Weakly-supervised Approach[C]//*IEEE International Conference on Computer Vision*. New York: IEEE Press, 2017: 398-407.
- [10] Bregler C, Malik J. Tracking People with Twists and Exponential Maps[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 1998: 8-15.
- [11] Taylor G W, Hinton G E, Roweis S T. Modeling Human Motion Using Binary Latent Variables[C]//*Advances in Neural Information Processing Systems*. Cambridge Massachusetts: MIT Press, 2007: 1345-1352.
- [12] Wan E A, van Der Merwe R. The Unscented Kalman Filter for Nonlinear Estimation[C]//*Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000*. New York: IEEE Press, 2000: 153-158.
- [13] van Der Merwe R. Sigma-point Kalman Filters for Probabilistic Inference in Dynamic State-space Models [D]. 2004.