

# 基于骨骼三维信息结合隐马尔科夫模型人体动作识别方法

隋龙飞<sup>1</sup>, 薛欢欢<sup>2</sup>, 翁政魁<sup>1</sup>

<sup>1</sup>嘉兴职业技术学院, 浙江 嘉兴

<sup>2</sup>嘉兴南湖学院, 浙江 嘉兴

收稿日期: 2022年1月20日; 录用日期: 2022年2月16日; 发布日期: 2022年2月23日

---

## 摘要

本文在人体动作描述采用关键骨骼点的深度三维坐标,通过各个关节的位置坐标关系来描述人体动作。进一步为每一类人体动作训练一个深度信息隐马尔科夫模型,从一定程度上解决了人体动作部分遮挡问题,对训练样本,参照得到的数据库完成编码,通过隐马尔科夫模型训练得到相应参数模型,通过训练得到的模型进行验证和测试,实验数据结果表明该方法具有精确率高的优点。

## 关键词

动作识别, 隐马尔科夫模型, 骨骼

---

# Human Motion Recognition Method Based on Bone 3D Information Combined with Hidden Markov Model

Longfei Sui<sup>1</sup>, Huanhuan Xue<sup>2</sup>, Zhengkui Weng<sup>1</sup>

<sup>1</sup>Jiaying Vocational and Technical College, Jiaying Zhejiang

<sup>2</sup>Jiaying Nanhu University, Jiaying Zhejiang

Received: Jan. 20<sup>th</sup>, 2022; accepted: Feb. 16<sup>th</sup>, 2022; published: Feb. 23<sup>rd</sup>, 2022

---

## Abstract

In this paper, the depth three-dimensional coordinates of key bone points are used in the description of human action, and the human action is described through the position coordinate rela-

tionship of each joint point. Further, a depth information hidden Markov model is trained for each kind of human action, which solves the problem of partial occlusion of human action to a certain extent. The training samples are encoded with reference to the obtained database, the corresponding parameter model is obtained through hidden Markov model training, and the model obtained through training is verified and tested. The experimental results show that this method has the advantage of high accuracy.

## Keywords

Motion Recognition, Hidden Markov Model, Bones

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

人体动作识别一直是计算机视觉、人工智能和模式识别等热门研究方向[1], 在人体动作识别的研究中, 需要准确地将人体动作信息通过技术手段转换成能够精确运算和分析的数据, 通过程序让计算机处理这些数据, 这些数据的获得就需要大量的人体动作数据作为支撑。传统的人体动作识别技术应用 OI。LSTM 神经网络分析算法、深度学习方法、CNN 算法等建立模型[2], 但在建立人体模型和程序算法的过程中, 运算准确率并不理想, 错误率较高。新加坡国立大学的 Shi Yuan Tang 等设计了一种可穿戴步态辅助系统。该系统通过脚腕佩戴力敏电阻来记录行走过程中的步态信息数据, 然后使用神经网络进行步态数据建模, 对步态数据进行分析和处理, 最终结果表明两种模型的验证精度均达到了 80% [3], 在融合多种数据进行识别方面, 毛峡等[4]提出一种基于 RGB 和骨骼信息的多模态人体行为识别框架。识别率依然不是很高, 还有很大的提升空间。微软的 Kinect 深度相机的上市, 使研究人员低成本获取视频中人体各关节点三维坐标信息和场景深度信息等[5] [6], 人体骨骼中关节的层次结构和不同语义角色为动作识别提供了重要信息[7] [8]。基于此本文提出了一种基于骨骼三维信息结合隐马尔科夫模型人体动作识别方法, 通过此方法来改进以上人体动作识别率低的问题。

## 2. 人体动作三维特征提取

人体在完成某一些动作时, 身体的各个关节空间坐标位置会发生一定的变化。本文使用 Kinect 深度相机获取人体骨骼关节点三维坐标作为实验特征变量, 利用关节点在不同动作位置变化来提取有用的动作特征, 以各关节点三维坐标的平均值作为坐标原点, 获取骨骼关节点在三维坐标系中的变化, 从而得到骨骼关节点空间位置特征。如图 1 所示本文获取 20 个人体骨骼。

首先获取坐标原点  $M_0 = (X_0, Y_0, Z_0)$ , 人体 20 个骨骼关节点的三维坐标的平均值即是坐标原点  $X_0 = (X_1 + X_2 + \dots + X_n) / 20$ ,  $Y_0 = (Y_1 + Y_2 + \dots + Y_n) / 20$ ,  $Z_0 = (Z_1 + Z_2 + \dots + Z_n) / 20$ ,  $M = M_i - M_0$  ( $i = 1, 2, \dots, 20$ ),  $M_i$  代表第  $i$  个骨骼点的三维坐标,  $M_0$  代表的是坐标原点。20 个关节点在 X、Y、Z 轴的变化量为  $M'_x = [\Delta X'_1, \Delta X'_2, \dots, \Delta X'_{20}]$ ,  $M'_y = [\Delta Y'_1, \Delta Y'_2, \dots, \Delta Y'_{20}]$ ,  $M'_z = [\Delta Z'_1, \Delta Z'_2, \dots, \Delta Z'_{20}]$ 。  $M_t = [M'_x, M'_y, M'_z]$  表示第  $t$  帧的特征向量。如果人体动作包含  $N$  帧  $M = [M_1, M_2, M_3, \dots, M_N]$ , 则每一帧得到的骨骼关节点向量的维度是  $20 \times 60$ 。为了减少人体高矮胖瘦对实验的影响, 对特征向量进行归一化, 同时除以头部  $M_1$  到坐标原点  $M_0$  的距离  $\vec{M} = M_1 - M_0$ 。

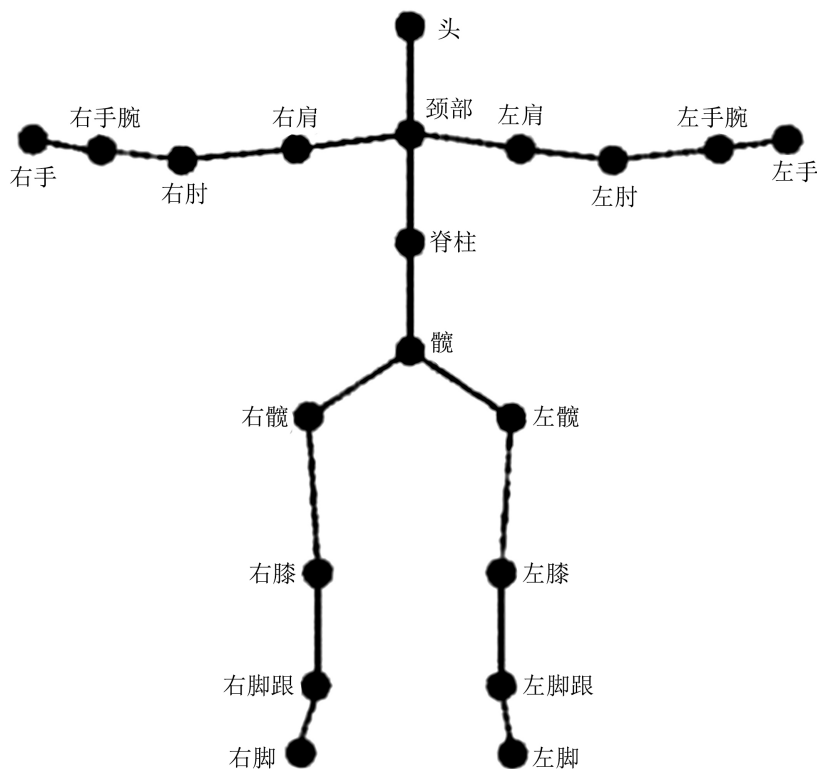


Figure 1. Three dimensional view of human skeleton  
图 1. 人体骨骼三维图

### 3. 基于隐马尔科夫的姿态表示

根据形成的骨架模型库，从视频中获得骨架三维信息，在实验中将训练视频提取出骨架之后，根据骨架信息库对其进行分类，从而获得骨架相应编号，这样就将人体动作的序列变成了可以使用隐马尔科夫模型性对应的观测值序列表示。

每个动作的隐马尔科夫模型由 5 各参数确定：

- 1)  $S = \{S_1, S_2, \dots, S_i, \dots, S_n\}$  代表模型的隐状态数，可以理解为动作的特征的动作数。
- 2)  $O = \{O_1, O_2, \dots, O_i, \dots, O_m\}$  观测值，代表骨架信息的种类数。
- 3)  $A = |S_{ji}|_{n \times n}$  是隐马尔可夫模型各隐状态之间的概率转换矩阵。
- 4)  $B = |S_i O_j|_{n \times m}$  是隐马尔可夫模型隐状态到观测状态之间的概率转换矩阵。
- 5)  $\pi = \{\pi_1, \pi_2, \dots, \pi_i, \dots, \pi_n\}$  为隐状态的初始概率矩阵。

从骨架信息库中得到的观测序列的编号，经过训练更新 HMM 与样本序列对应的初始参数，这就是 HMM 对这个动作的基本描述，待识别的动作序列，经过得到的动作编号序列，可通过个续联的模型，就是更新之后的 HMM 新参数，来完成每个模型的匹配值计算。

#### HMM 模型的训练方法

通过 EM 算法，重复迭代完成模型的训练。主要包括以下 4 个步骤。

- 1) 初始化参数：选定合适的隐状态个数  $S = \{S_1, S_2, \dots, S_i, \dots, S_n\}$ ，根据骨架信息库确定观测序列  $O = \{O_1, O_2, \dots, O_i, \dots, O_m\}$ ，随机生成  $\lambda = (A, B, \pi)$ 。
- 2) 确定迭代次数，通常迭代次数可取 10 次左右。

3) 当前模型相关概率的计算: 按照状态  $i_t$  的观测概率分布  $b_{i_t}(k)$  生成  $o_t$ , 按照状态  $i_t$  的状态转移概率分布  $\{a_{i_t i_{t+1}}\}$  生成状态  $i_{t+1}$ ,  $i_{t+1} = 1, 2, 3, \dots, 20$ 。

4) 初始化概率分布矩阵:  $\pi = \{\pi_1, \pi_2, \dots, \pi_i, \dots, \pi_n\}$ 。

HMM 的评估问题利用前向或者后向概率计算方法来解决, 前向概率计算由前向算法公式递归关系计算, 后向概率计算由后向算法公式递归关系计算。

前向算法:

1) 初始值:

$$a_1(i) = \pi_i b_i(o_1)$$

2) 递归对  $t = 1, 2, \dots, T - 1$

$$a_{t+1}(i) = \left[ \sum_{j=1}^N a_t(j) a_{ji} \right] b_i(o_{t+1}), \quad i = 1, 2, \dots, N$$

3) 终止

$$P(O|\lambda) = \sum_{i=1}^N a_T(i)$$

## 4. 实验与分析

### 4.1. 训练和验证模型

隐马尔科夫模型的参数设置: 隐状态数  $S = 20$ , 可观测值  $O = 30$ , 状态转移矩阵  $A = |S_{ji}|_{20 \times 20}$ , 观测矩阵  $B = |S_i O_j|_{20 \times 30}$ , 初始概率  $\pi_i = \frac{1}{20}$ ,  $i = 1, 2, \dots, 20$ 。

实验对 5 种常见动作走路、跑步、坐下、站立、下蹲进行测试。实验中, 选取 30 个身高和体型上有差异的测试人员, 每个人重复做动作 30 次, 每组动作具有 900 个动作样本, 共有 4500 个动作样本。本文对 5 种动作走路、跑步、坐下、站立、下蹲做实验, 实验结果如下表 1 所示。

**Table 1.** Confusion matrix of human motion recognition in this paper

**表 1.** 本文人体动作识别混淆矩阵

输入	识别率(%)					时间(s)
	走路	跑步	坐下	站立	下蹲	
走路	92.8	4.2	3.0	0	0	0.31
跑步	4.3	94.8	0.9	0	0	0.35
坐下	3.2	0.8	95.6	0	0.4	0.36
站立	0	0	0	100	0	0.42
下蹲	0	4.8	1.6	0	93.6	0.32

基于传统 K-means 和 DTW 算法的人体动作识别结果, 如下表 2 所示。

**Table 2.** Human motion recognition confusion matrix of traditional K-means**表 2.** 传统 K-means 的人体动作识别混淆矩阵

输入	识别率(%)					时间(s)
	走路	跑步	坐下	站立	下蹲	
走路	90.8	2.0	2.0	2.3	2.9	0.53
跑步	0.8	91.8	3.6	3.5	0.3	0.61
坐下	2.5	3.2	89.6	2.2	2.4	0.59
站立	2.0	1	5.2	90.6	1.2	0.66
下蹲	6.2	1.1	1.1	0	91.6	0.52

通过表 1 和表 2 可以看出, 本文的人体动作识别算法, 对以上 5 中人体动作识别的准确率有了明显的提高, 且识别速度比之前也有明显的加快。

#### 4.2. 方法优化

为了提高识别的鲁棒性, 对动作人体进行关键帧提取, 每个动作用 4 个关键动作进行表示, 对相似动作进行合并, 共得到 15 个关键帧, 得到新的隐马尔科夫模型参数值  $S = 10$ ,  $O = 15$ , 对 Weizman 数据库中的走路、跑步、跳跃、站立、坐下进行测试得到的结果如表 3 所示。

**Table 3.** Confusion matrix of human motion recognition in this paper**表 3.** 本文人体动作识别混淆矩阵

输入	识别率(%)					时间(s)
	走路	跑步	坐下	站立	下蹲	
走路	93.8	4.0	2.2	0	0	0.30
跑步	4.1	94.9	1.0	0	0	0.34
坐下	4.2	0	95.8	0	0	0.32
站立	0	0	0	100	0	0.29
下蹲	3.6	0.6	0.2	0	95.6	0.38

#### 4.3. 算法比较

在实验过程中, 本文的实验与其他两种动作识别方法进行了对比。方法一基于多模态的人体动作识别研究[9], 方法二文献[10]基于关节角度特征向量, 使用 SVM 分类器的动作识别方法。表 4 为三种方法在识别率上的混淆矩阵对比。从表 4 的结果可以看出: 方法一的平均识别率为 90.18%, 方法二的平均识别率为 88.54%, 容易受到噪声干扰。本文方法的平均识别率为 96.00%, 明显的高于前两种方法的识别率, 同时本文方法克服了人体骨骼点空间位置和角度的变化, 及被测人员体型差异的影响, 如下表 4 所示。

**Table 4.** Algorithm comparison  
**表 4.** 算法比较

(a) 方法一的识别率混淆矩阵

输入	识别率(%)				
	走路	跑步	坐下	站立	下蹲
走路	90.8	3.2	2.4	3.6	0
跑步	3.1	89.9	2.1	0	4.9
坐下	1.2	0	90.5	5.2	3.1
站立	3.4	2.2	1.0	90	3.4
下蹲	3.0	2.6	3.2	0.6	90.6

(b) 方法二的识别率混淆矩阵

输入	识别率(%)				
	走路	跑步	坐下	站立	下蹲
走路	86.2	1.2	3.1	4.5	5.0
跑步	5.4	87.1	1.0	3	3.5
坐下	3.1	3.3	88.8	2.5	2.3
站立	4	1.4	1.5	90	3.1
下蹲	1.2	1.0	3.8	3.4	90.6

(c) 本文的识别率混淆矩阵

输入	识别率(%)				
	走路	跑步	坐下	站立	下蹲
走路	93.9	4.1	2.0	0	0
跑步	3.2	94.8	1.2	0.8	0
坐下	0	0	95.7	0	4.3
站立	0	0	0	100	0
下蹲	0	0	4.4	0	95.6

## 5. 结语

基于关键骨骼点深度三维坐标和隐马尔科夫模型提出了一种新的人体动作识别算法, 对不同用户的 5 种基本生活动作可达到 96% 的识别率, 是一种可靠的人体动作识别方案, 该方法可以用于智能监控识别系统, 为公安系统提供人员动作数据, 降低犯罪率。

## 基金项目

浙江省教育厅, 校企合作访问工程师项目, FG2020272, 基于多模块融合的人体异常行为智能识别

技术研究；浙江省自然科学基金委员会，公益技术研究计划，LGF21F020007，智能视频监控中基于情境感知和多模态的人体异常行为识别技术研究。

## 参考文献

- [1] 高常强, 顾桂梅, 赵建龙, 等. 基于单幅光栅条纹图像的钢轨表面三维重构研究[J]. 传感器与微系统, 2019, 38(7): 57-60.
- [2] 毕雪超. 基于计算机视觉的舞蹈视频动作识别技术研究[J]. 电子设计工程, 2020, 28(7): 151-155.
- [3] Tang, S.Y., Hoang, N.S., Chui, C.K., *et al.* (2019) Development of Wearable Gait Assistive Device Using Recurrent Neural Network. 2019 *IEEE/SICE International Symposium on System Integration (SII)*, Paris, 14-16 January 2019, 626-631. <https://doi.org/10.1109/SII.2019.8700415>
- [4] 毛峡, 王岚, 李建军. 一种基于 RGB-D 特征融合的人体行为识别框架[J]. 计算机科学, 2018, 45(8): 22-27.
- [5] 刘艳, 张伟, 厉健晖. 基于消隐点的摄像机标定关键技术研究现状及发展[J]. 传感器与微系统, 2018, 37(3): 3-6, 13.
- [6] 黄晴晴, 周风余, 刘美珍. 基于视频的人体动作识别综述[J]. 计算机应用研究, 2020, 37(11): 3213-3219.
- [7] 叶松涛, 周扬正, 范红杰, 陈正雷. 融合因果关系和时空图卷积网络的人体动作识别[J]. 计算机科学, 2021(z2): 130-135.
- [8] 张良, 钱毅敏. 基于深度图像和骨骼信息的人体动作识别方法[J]. 中国民航大学学报, 2021, 39(2): 54-60.
- [9] 丁川, 崔爱浩, 齐林, 帖云. 基于多模态的人体动作识别研究[J]. 人工智能, 2021(1): 50-58.
- [10] 梅阳, 王永雄, 秦琪. 一种基于关键帧的人体行为识别方法[J]. 光学技术, 2017, 43(4): 323-328.