

A Skeleton Adaptive Human Motion Similarity Calculation Method

Shengnan Li^{1, a}, Yan Ma^{1, b*}, Hui Huang^{1, c} and Shunbao Li^{1, d}

¹Collage of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai, China

^alsn_199310@163.com, ^bma-yan@shnu.edu.cn, ^chuanghui@shnu.edu.cn, ^dlsb@shnu.edu.cn

*The corresponding author

Abstract. Human motion similarity algorithm is important for human behavior recognition. However, the accuracy and stability of motion similarity algorithm will be affected by different data acquisition objects and object actions. In order to address this issue, a skeleton-adaptive human motion similarity method is proposed. First, the data of human joint points is acquired by Kinect. Second, the joints of the motion to be tested are corrected with the length of template skeleton. Third, the weight of joints are adjusted adaptively. Finally, the similarity between motions is calculated. The experimental results show that the performance of the proposed method is more stable and accurate.

Keywords: Human Behavior Recognition; Kinect; Joint Position Correction; Weight Adaptation; Motion Similarity

一种骨架自适应的人体动作相似度计算方法

李盛楠¹, 马燕¹, 黄慧¹, 李顺宝¹

(1. 上海师范大学 信息与机电工程学院、计算机系, 上海市 200234)

摘要: 人体动作相似度算法对人体行为识别至关重要。但数据采集对象、对象动作的不同都会影响动作相似度算法的准确度和稳定性。为解决该问题, 提出一种骨架自适应的人体动作相似度方法。首先, 用 Kinect 获取人体关节点数据; 其次, 利用模板骨架长度对待测试动作的关节点位置修正; 然后, 关节点权重自适应调整; 最后, 计算动作间的动作相似度。实验结果表明, 该方法的动作相似度结果数值表现更稳定且准确度更高。

关键词: 人体行为识别; Kinect; 关节点位置修正; 权重自适应; 动作相似度

中图分类号: TP 391.4 文献标志码: A

引言

人体动作相似度研究是人体行为识别[1]的重要研究分支之一, 研究内容是从一些未知信息中自动分析并识别行为, 其研究成果广泛应用于虚拟现实、运动行为分析等领域。

一个好的人体动作相似度算法的设计应注重三个方面: ①数据易于收集且占用空间小; ②算法识别结果准确度高; ③算法运行时间短。以前的研究中, 研究者主要用二维图像[2-5]信息, 二维图像和视频在光照和阴影的影响下, 易使目标识别不稳定。

微软公司推出的 Kinect 体感设备[6], 具有卓越的深度感知和人体关节点识别能力, 可以有效避免光照等环境因素的影响。为人体行为识别研究提供了更好的数据来源。Kinect 近年来被广泛应用于人体行为

识别领域[7-10]。Thanh[11]引入图像算法结合 Kinect 数据进行动作识别，论文中只收集局部数据，数据上面有局限性。Girshick[12]利用 Kinect 深度数据研究人体关节行为变化，直接对原始深度图像回归，而原始数据不可避免脏数据的问题。

鉴于这些问题，本算法应运而生。做了如下工作：①提出人体关节映射算法；②提出了权值自适应的骨架相似度算法。

1 数据采集

采集数据来源是实验室志愿者，采集 6 人，每人 10 个动作，共计 60 组动作数据。采用微软 Kinect 体感设备，利用深度信息，可以准确识别人体骨骼关节点。关节点共 25 处，每处详细编号和位置见图 1：

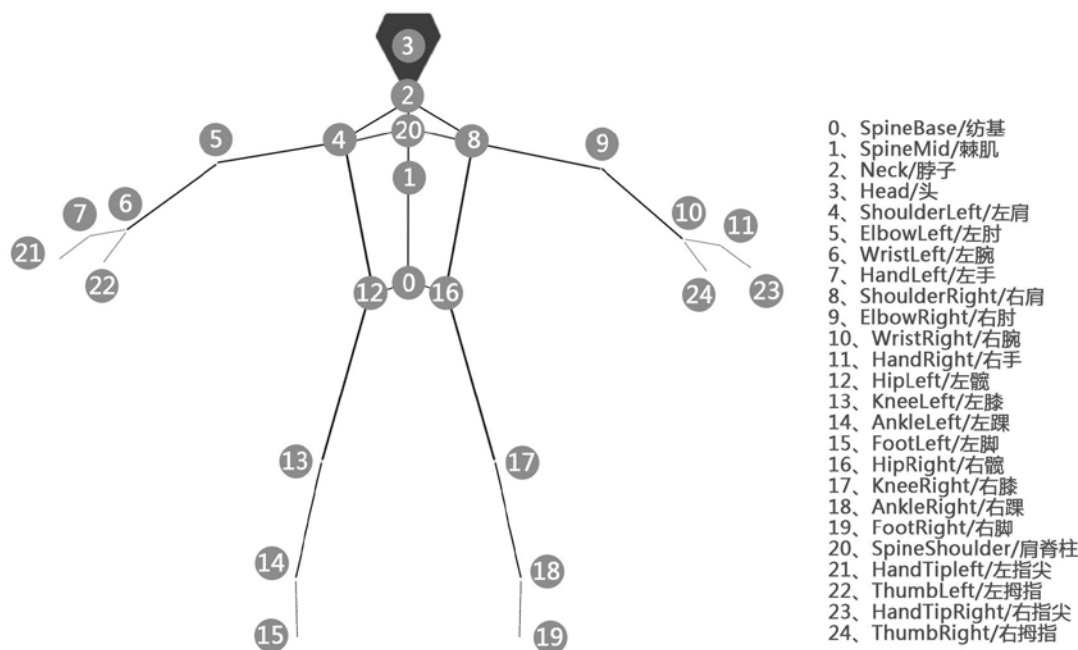


图 1 Kinect 人体关节点

Kinect 获取的关节点信息是深度图像数据，先转换到三维空间坐标。转换过程可以调用 Kinect for Windows SDK 提供的映射函数，在此不详述。关节点信息记为 J ，表示为： $J = (j_0, j_1, j_2, \dots, j_{23}, j_{24})$ 。每个关节点的位置信息包括三个维度，记为 j_i ，表示为： $j_i = (x_i, y_i, z_i)$ ， $i \in [0, 24]$ 。模板动作信息的 25 个关节点记为 t ，表示为： $t = (j_0^t, j_1^t, j_2^t, \dots, j_{23}^t, j_{24}^t)$ 。待检测动作信息的 25 个关节点，记为 a ，表示为： $a = (j_0^a, j_1^a, j_2^a, \dots, j_{23}^a, j_{24}^a)$ 。

2 动作相似度算法

由于人类躯体比例会有不同，比如模板信息是成年人，而待检测目标是小孩或其他极端情况。为了保证算法的通用性，在动作相似度计算前，将待检测骨架长度信息映射到模板骨架长度，然后进行坐标平移，最后进行相似度计算。

2.1 坐标变换

根据关节点算出模板动作 t 和待检测动作 a 的各骨架长度，然后将 a 骨架长度变换到 t 的对应骨架长

度。 a 骨架长度变换后, 选取 t 和 a 的基准点, 以基准点重合的方式将 a 平移到目标位置。具体变换过程如下:

- ① 规划骨架线段, 25 个关节点共有 24 条骨架线段。因骨架线段起始点和结束点的顺序对算法结果有影响, 所以每条线段的起始点和结束点是固定的。以编号 1 的关节点为总起点, 到达 23 号关节点的路径为: 1-20-8-9-10-11-23。表示这些线段时, 1 是起点 20 是结束点, 20 是起始点 8 是结束点, 同理确定后面线段起始点和结束点。每条线段起始关节点的编号记为 s , 结束关节点的编号记为 e , k 代表第 k 条线段, 长度记为 $l_k(j_s, j_e)$, $k \in [0, 23]$, $s \in [0, 24]$, $e \in [0, 24]$, 骨架线段起点和结束点的详细信息见表 1。

表 1 骨架各线段起始、终止点的编号

k	s	e	k	s	e	k	s	e
0	2	3	8	7	21	16	0	12
1	20	2	9	8	9	17	12	13
2	20	4	10	9	10	18	13	14
3	20	8	11	10	11	19	14	15
4	4	5	12	10	24	20	0	16
5	5	6	13	11	23	21	16	17
6	6	7	14	1	20	22	17	18
7	6	22	15	1	0	23	18	19

以上 24 条线段, 一起记为 L : $L = \{l_0(j_2, j_3), l_1(j_{20}, j_2), \dots, l_{22}(j_{17}, j_{18}), l_{23}(j_{18}, j_{19})\}$ 。

t 的骨架线段记为 L_t : $L_t = \{l_0^t(j_2^t, j_3^t), l_1^t(j_{20}^t, j_2^t), \dots, l_{22}^t(j_{17}^t, j_{18}^t), l_{23}^t(j_{18}^t, j_{19}^t)\}$ 。

a 的骨架线段记为 L_a : $L_a = \{l_0^a(j_2^a, j_3^a), l_1^a(j_{20}^a, j_2^a), \dots, l_{22}^a(j_{17}^a, j_{18}^a), l_{23}^a(j_{18}^a, j_{19}^a)\}$ 。

- ② 计算骨架各线段长度 $l_k(j_s, j_e)$, $k \in [0, 23]$, $s \in [0, 24]$, $e \in [0, 24]$, 计算公式:

$$l_k(j_s, j_e) = \sqrt{(x_e^t - x_s^t)^2 + (y_e^t - y_s^t)^2 + (z_e^t - z_s^t)^2}$$

- ③ 计算 a 骨架线段结束点 a_e 需要变换的三维坐标数据记为 M_e , 详细过程:

$$M_e = (x_e^M, y_e^M, z_e^M), e \in [0, 24]$$

$$\sigma = \frac{l_k^t(j_s^t, j_e^t)}{l_k^a(j_s^a, j_e^a)} - 1, k \in [0, 23], s \in [0, 24]$$

$$x_e^M = (x_e^t - x_s^t) * \sigma$$

$$y_e^M = (y_e^t - y_s^t) * \sigma$$

$$z_e^M = (z_e^t - z_s^t) * \sigma$$

- ④ 变换 a 中各个关节点坐标。变换关节点的目的是为了匹配骨架线段的长度, 所以关节点变换时应考虑

到线段的起始点位置、结束点位置和线段具体位置。为了计算效率和算法简洁，选用脊椎点 j_1 作为基点，其他关节点均是以该点变换相对位置。结合上一步计算的每个关节点需要变换的数据，以关节点 j_{20} 和 j_{21} 的坐标变换为例，参照图 1 做出一些说明。 j_{20} 与 j_1 直连，因此 a 中关节点 j_{20}^a 的坐标变换过程是： $j_{20}^a = j_{20}^a + M_{20}$ 。关节点 j_1 到达 j_{21} 的路径为 $j_1 \rightarrow j_{20} \rightarrow j_4 \rightarrow j_5 \rightarrow j_6 \rightarrow j_7 \rightarrow j_{21}$ ， $j_1 \rightarrow j_{20}$ 段的坐标变换参照上文， $j_{20} \rightarrow j_4$ 段的坐标变换应该 $j_1 \rightarrow j_{20}$ 段的基础上进行，那么， $j_1 \rightarrow j_{21}$ 的变换和六段骨架线段有关系，因此 a 中关节点 j_{22}^a 的坐标变换过程是： $j_{21}^a = j_{22}^a + M_{20} + M_4 + M_5 + M_6 + M_7 + M_{21}$ 。

下列出 j_1 到所有关节点（不包含 j_1 ）的路径，表中的“点”栏代表从 j_1 出发到达的关节点编号，“路径”栏是从 j_1 出发到达目标关节点所经过关节点的顺序编号，详见表 2。

表 2 关节点 j_1 到其他关节点的路径

点	路径	点	路径	点	路径
0	1-0	17	1-0-16-17	7	1-20-4-5-6-7
20	1-20	3	1-20-2-3	15	1-0-12-13-14-15
12	1-0-12	5	1-20-4-5	19	1-0-16-17-18-19
16	1-0-16	9	1-20-8-9	11	1-20-8-9-10-11
4	1-20-4	14	1-0-12-13-14	21	1-20-4-5-6-7-21
8	1-20-8	18	1-0-16-17-18	22	1-20-4-5-6-22
2	1-20-2	6	1-20-4-5-6	24	1-20-8-9-10-24
13	1-0-12-13	10	1-20-8-9-10	23	1-20-8-9-10-11-23

- ⑤ a 骨架长度变换后，只能保证骨架线段的比例正确，不能保证相对位置正确，所以还需要对平移。选取 t 中的脊椎点 j_1^t 作为基准点，使 a 中的脊椎点 j_1^a 平移到 j_1^t 位置， a 中的其他关节点做和 j_1^a 平移路径一样的平移动作。平移过程： $j_p^a = j_p^a + (j_1^t - j_1^a)$ ， $p \in [0,24]$ 。

2.2 相似度计算

由于人类骨架各段长度不一致，在计算相似度时，每个关节点的权重应该是不一样的。因此提出了权值自适应的骨架相似度算法。具体过程：

- ① 定义关节点的权重 W ，25 个关节点对应 25 个权重： $W = (w_0, w_1, w_2, \dots, w_{23}, w_{24})$ 。
- ② 权重自适应取值。取值原理是模板信息中较长骨骼边的尾关节点相应权值较大，相应的较短骨骼边的尾关节点权值较小，基准点 j_1 对应的 w_1 为 0，具体见：

$$w_e = \frac{l_k^t(j_s^t, j_e^t)}{\sum_{d=0}^{23} l_d^t(j_s^t, j_e^t)}, k \in [0,23], s \in [0,24], e \in [0,24]$$

- ③ 计算动作相似度。先计算 t 和 a 中对应关节点的距离 $d_q(j_q^t, j_q^a)$ ：

$$d_q(j_q^t, j_q^a) = \sqrt{(x_q^t - x_q^a)^2 + (y_q^t - y_q^a)^2 + (z_q^t - z_q^a)^2}, \quad q \in [0,24]$$

再将各个距离按照以上权重计算汇总距离 $D(t, a)$:

$$D(t, a) = \sum_{q=0}^{24} ((d_q(j_q^t, j_q^a) * w_q)$$

D 是一个大于等于 0 的数, 且数值越大, 对应的动作相似度越小。参考 *Sigmoid* 函数, ε 是预设 *Kinect* 误差因子, 距离 $D(t, a)$ 和相似度 $S(t, a)$ 的关系:

$$S(t, a) = \left(\frac{1}{1 + e^{D(t, a)/\varepsilon}} \right) * 2, D \in (0, +\infty), S \in (0, 1)$$

至此, 求出测试动作和模板动作的相似度。

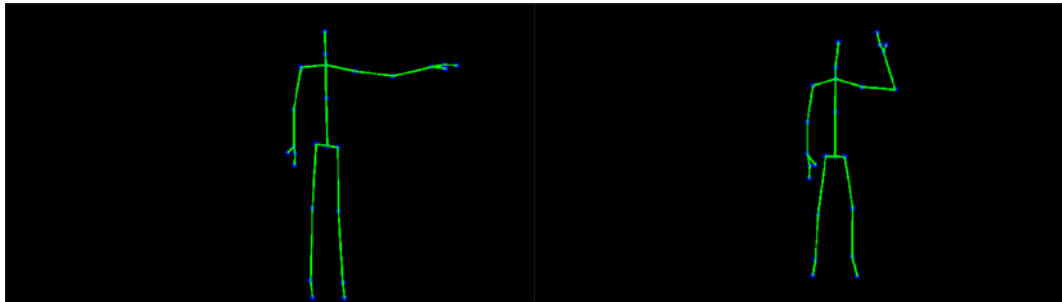
2.3 动作相似度结果分析

以图 2 中(a)和(b)分别为模板信息和待检测信息来分析。图 2 中(c)是模板信息的骨架图, 图 2 中(d)是待检测信息的骨架图, 图 2 中(e)是待检测关节点经过坐标变换后的骨架图, 图 2 中(f)是坐标变换后平移的骨架点图, 详见图 2。



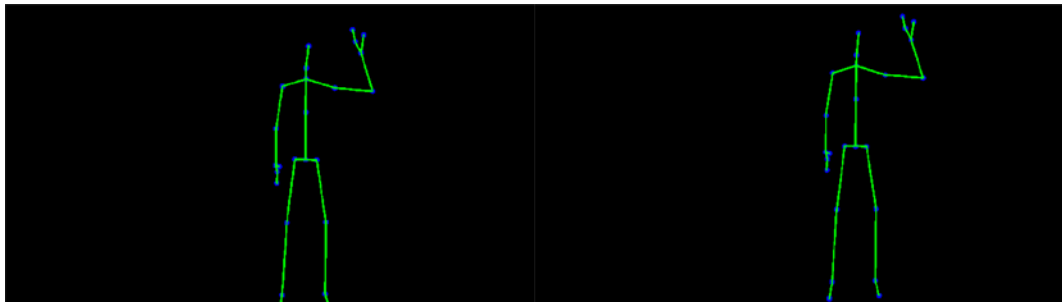
(a) 模板图

(b) 待检测图



(c) 模板骨架图

(d) 待测试骨架图



(e) 待测试坐标变换后

(f) 待测试平移后图

图 2 相似度计算过程的骨架信息

实验中,自适应的 25 个权值 $w_0 \sim w_{24}$ 分别是:0.0705583、0、0.0174371、0.0402958、0.0375586、0.058309、0.0499353、0.0200813、0.0432484、0.0528944、0.0628691、0.0117717、0.014971、0.0930125、0.0706322、0.0259879、0.0154678、0.0825626、0.0754748、0.0324475、0.0525174、0.0186768、0.0217738、0.0202465、0.0112703。误差因子 ε 选用 3000,不加权值计算时,动作相似度结果是 75.77%,经过自适应权值计算后,动作相似度结果是 92.41%,更加符合志愿者观感,与该待检测人自身体验感觉的相似度更加吻合。

3 结语

在人体关节映射算法中,以模板信息中人体各个关节为基础,对待测试信息进行处理,有效解决了不同体型之间的动作相似度计算。选取了 30 组动作信息,经过实验测试, Kinect 数据经过关节映射算法处理后,在不加权值自适应的骨架相似度算法时,所有动作之间的距离值显著减少,相比关节映射算法处理前,组别之间距离值的方差减少 8.2%,动作相似度结果平均高了 14.31%。经过数据回溯调查,发现 Kinect 收集数据有是空数据,设备会自动把数据置 0,这样计算来的动作之间距离值显著增大。人体关节映射算法能够修正该种“脏”数据,使同种动作的相似度提高很多,给行为相似度的求解提供了一个优秀的前置处理方法。

在权值自适应的骨架相似度算法中,以模板信息中人体各个线段为参考,在计算动作距离值时给每个关节增加一个权值。针对前述经过关节映射算法处理后的数据,该算法,加比不加动作相似度结果平均高了 4.1%。经过数据回溯调查,相似度变高的原因在于人体长骨架主要在下半身,相对比较稳定和好调整。对于上半身来说,骨架线段长度相对较短,而且行为上不好准确把握,权值自适应的骨架相似度算法能够良好的修正人体该种情况。

综合所有实验数据,同时用人体关节映射算法和权值自适应骨架相似度算法,该次实验 ε 值设置为 3000,同一种动作间的动作相似度结果平均高了 14.89%,平均动作相似度值是 95.4%,所以该动作相似度方法效果良好。

Acknowledgement

The authors are grateful to the support of the National Natural Science Foundation of China (No.61373004, No.61501297).

References

- [1] Suraj Prakash Sahoo, Samit Ari. On an algorithm for human action recognition[J]. Expert Systems with Applications, 2019, 115: 524-534.
- [2] Maity, Satyabrata, Bhattacharjee, et al. A Novel Approach for Human Action Recognition from Silhouette Images[J]. IETE Journal of Research, 2017, 63(2): 160-171.
- [3] Beaudry, Cyrille¹, Péteri, et al. An efficient and sparse approach for large scale human action recognition in videos[J]. Machine Vision and Applications, 2016, 27(4): 529-543.
- [4] Mona, Elsayed, et al. An enhanced method for human action recognition[J]. Journal of Advanced Research, 2015, 6(2): 163-169.

- [5] Asteriadis, Stylianos, Daras, et al. Landmark-based multimodal human action recognition [J]. Multimedia Tools and Applications, 2017, 76(3): 4505-4521.
- [6] Xu Zongyi, Zhang Qianni, Cheng Shiyang. Multilevel active registration for kinect human body scans: from low quality to high quality[J]. Multimedia Systems, 2018, 24(3): 257-270.
- [7] Seddik, Gazzah. Human-action recognition using a multi-layered fusion scheme of Kinect modalities[J]. Iet Computer Vision, 2017, 11(7):530-540.
- [8] Liu Xueping, Li Yibo, Li Xiaoming, et al. Multi-feature consultation model for human action recognition in depth video sequence[J]. The Journal of Engineering. 2018, (16): 1498-1502.
- [9] Gavrilova, Wang Yingxu, Ahmed, et al. Kinect Sensor Gesture and Activity Recognition: New Applications for Consumer Cognitive Systems[J]. IEEE Consumer Electronics Magazine, 2017, 7(1): 88-94.
- [10] Li Qiannan, Wang Yafang, Sharf, et al. Classification of gait anomalies from Kinect[J]. Visual Computer, 2018, 34(2): 229-241.
- [11] Thanh, Chen Fan, Kotani, et al. Extraction of Discriminative Patterns from Skeleton Sequences for Human Action Recognition[C]//IEEE RIVF International Conference on Computing and Communication Technologies, Research, Innovation, and Vision for the Future. IEEE, 2014:1-6.
- [12] Girshick R, Shotton J, Kohli P, et al. Efficient regression of general-activity human poses from depth images[C]// International Conference on Computer Vision. IEEE Computer Society, 2011:415-422.