



引用格式:胡新荣,王梦鸽,刘军平,等.基于Kinect的人体三维动作实时动态识别[J].科学技术与工程,2020,20(34):14133-14138

Hu Xinrong, Wang Mengge, Liu Junping, et al. Real-time dynamic recognition of human three-dimensional posture based on Kinect[J]. Science Technology and Engineering, 2020, 20(34): 14133-14138

# 基于 Kinect 的人体三维动作实时动态识别

胡新荣<sup>1,2</sup>, 王梦鸽<sup>1,2</sup>, 刘军平<sup>1,2\*</sup>, 彭 涛<sup>1,2</sup>

(1. 湖北省服装信息化工程技术研究中心, 武汉 430200; 2. 武汉纺织大学数学与计算机学院, 武汉 430200)

**摘要** 针对传统的二维图片动作识别算法识别率相对不高、实时性不强的问题,提出一种三维的人体动作实时识别的方法。该方法首先通过 Kinect 获取人体三维骨骼数据,然后对骨骼数据信息采取归一化的方法进行数据对齐的预处理,使得与实时数据与标准数据的角度阈值和距离参考值统一。最后与标准动作采用多特征融合的识别算法对动作进行识别与匹配的方法,并在此基础上改进基于关节点角度的动作识别方法。实验结果表明,本文方法运行速度较快,可有效消除角度测量不稳定以及距离测量无法检测方向上的差异造成动作匹配不准确。满足三维动作识别的实时性、鲁棒性要求。

**关键词** 三维; 动作识别; 实时识别; 骨骼数据; 动作匹配

中图法分类号 TP391.41; 文献标志码 A

## Real-time Dynamic Recognition of Human Three-dimensional Posture Based on Kinect

HU Xin-rong<sup>1,2</sup>, WANG Meng-ge<sup>1,2</sup>, LIU Jun-ping<sup>1,2\*</sup>, PENG Tao<sup>1,2</sup>

(1. Engineering Research Center, Hubei Province for Clothing Information, Wuhan 430200, China;

2. Mathematics and Computer School, Wuhan Textile University, Wuhan 430200, China)

**[Abstract]** Aiming at the problems of low recognition rate and low real-time performance of the traditional two-dimensional image motion recognition algorithm, a three-dimensional real-time human motion recognition method was proposed. Firstly, Kinect was used to obtain the three-dimensional bone data of human body. Then, the normalization method was used to preprocess the data alignment. Finally, the recognition algorithm of multi feature fusion was used to recognize and match the standard actions. The experimental results show that the method runs fast and can effectively eliminate the movement recognition inaccuracy caused by the instability of angle measurement and the inability of distance measurement to quantitatively describe the movement difference. It can meet the real-time and robust requirements of 3D motion recognition.

**[Key words]** 三 维; 动 作 识 别; 实 时 识 别; 骨 骼 数 据; 动 作 匹 配

随着人机交互和计算机技术的发展,人的身体作为人机交互的重要媒介之一,得到越来越多的研究人员的关注。人体动作更是能够表现出非常多的信息,在各个领域都发挥着重要的作用。因此,越来越多的人员投入到有关人体动作识别的研究。

中外很多机构都在做相关的研究,这些研究方法可分为三类。第一类是使用骨架模型进行动作识别;第二类是直接使用深度图像进行动作识别。第三类是同时使用骨骼关节点和深度图像进行动作识别。其中,Cao 等<sup>[1]</sup>和 Chen 等<sup>[2]</sup>使用深度学习方法的人体姿态识别可以达到很高的准确性。Fang 等<sup>[3]</sup>识别人体姿态使用的是 Stacked Hourglass 方

法,致力于解决检测带来的定位误差,但是多人识别速度较慢。周彤彤等<sup>[4]</sup>采用可穿戴式传感器(头盔、手套、加速传感器等)进行人体动作识别,相对准确,实时性高。但是,一方面设备需要加大成本,另一方面穿戴规范有一定的要求,造成方便性降低和准确性降低。蒲兴成等<sup>[5]</sup>采用改进的 Hu 矩阵算法,达到了很高的识别率,但是采用这种方法不能满足对系统的实时动作匹配。

利用 Kinect 进行人体姿态识别的研究众多<sup>[6-15]</sup>。例如,战荫伟等<sup>[7]</sup>采用角度测量的方法进行动作识别,由于没有对骨骼的三维数据实行预先分类,导致关键动作提取不够准确;李红波等<sup>[9]</sup>使

收稿日期: 2019-12-19; 修订日期: 2020-05-31

基金项目: 湖北省高等学校优秀中青年科技创新团队计划(T201807)

第一作者: 胡新荣(1973—),女,汉族,湖北武汉人,博士,教授。研究方向: 计算机图形图像处理,虚拟现实,计算机视觉,可视化计算。

E-mail:921498291@qq.com。

\*通信作者: 刘军平(1980—),男,汉族,博士,讲师。研究方向:工业大数据、人工智能。E-mail:7710981@qq.com。

用特征矢量的方法对动作进行匹配,引入改进的动作识别评分公式,有较好的识别率和实时性,但是使用的方法只是对准静态的人体动作进行匹配,难以对动态动作进行实时识别;吴剑峰等<sup>[15]</sup>使用两点法测量三维姿势关节点,使用隐马尔可夫模型也达到了很好的效果,但是仅有角度作为动作测量的标准,一定程度上由于角度不稳定,从而导致数据特征测量存在误差。在使用的算法比较中<sup>[16-18]</sup>,由于矢量特征在对目标多从方向上区分差异,而对数值缺乏敏感性,对动作的判定有一定的误差。使用深度图像进行动作识别的方法通常是提取点云特征或表面法矢量特征进行动作识别<sup>[19-20]</sup>,增加一定的计算成本。

为了提高动作识别的准确度和测量的实时性以及可行性,提出一种人体多特征融合的识别算法,以期提高动作匹配的精确度,配合人物动作的动态匹配,达到较高的实时性。

## 1 骨骼数据处理

关节点在三维空间中位置信息采用 Kinect 获取。Kinect 所拥有的骨骼追踪技术,可以实时精确地获取到人体 25 个关节点位置的三维坐标。根据关节位置的三维信息可以表示人体各个关节点组成的角度特征和距离特征。如图 1 所示为 Kinect 提取的人体骨架中的 25 个关节点位置信息。

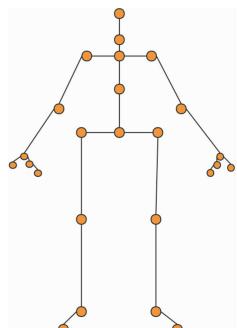


图 1 人体骨骼关键点

Fig. 1 Key points of human skeleton

### 1.1 归一化处理

经过实验表明,将获取的三维数据进行标准化处理,将标准动作特征和实时动作特征进行对齐处理,可以提高计算速度和精度。

在涉及使用距离来度量相似度时,使用线性函数标准化会产生误差,因此使用标准差归一化处理。选取每 10 ms 内的数据寻找一组关键动作进行匹配。运用式(1)对数据进行归一化处理。

$$A_{\text{norm1}} = (A - \mu) / \sigma \quad (1)$$

式(1)中: $\mu$  和  $\sigma$  分别为 10 ms 内获取的三维数据的

均值和方差;  $A_{\text{norm1}}$  为标准动作归一化处理后的结果。

在涉及角度特征度量相似度时,使用线性函数标准化的方式,选取每 10 ms 内的数据,寻找一组关键动作进行匹配。运用式(2)对数据进行归一化处理。

$$A_{\text{norm2}} = \frac{A - A_{\min}}{A_{\max} - A_{\min}} \quad (2)$$

式(2)中:  $A_{\max}$  和  $A_{\min}$  分别为 10 ms 内获取的三维数据的最大值和最小值;  $A_{\text{norm2}}$  为三维数据动作归一化处理后的结果。

### 1.2 关节点相对距离特征

距离特征常用文本相似度测量<sup>[22]</sup>,引入距离相似度度量人体三维动作。由于各个关节点之间的距离不尽相同,它们对动作匹配的贡献度也不同。欧几里得距离度量会受指标不同单位刻度的影响,所以需要先进行标准化。假设经过标准差归一化处理后的标准动作和实时动作中对应关节点之间位置特征点表示为  $(x_b, y_b, z_b)$  和  $(x_s, y_s, z_s)$ , 对应的距离为

$$d = \sqrt{(x_b - x_s)^2 + (y_b - y_s)^2 + (z_b - z_s)^2} \quad (3)$$

通过式(3),获取标准骨架与实时骨架中 25 个关节点的相对距离描述符  $P_1, P_1 = \{d_1, d_2, \dots, d_{25}\}$ 。

根据各矢量可以得到一个归一化的距离值( $D_1$ )。 $D_1$ 越大,则表明标准动作和实时动作的差异性越大:

$$D_1 = \sum_{i=1}^{25} d_i, \quad i = 1, 2, \dots, 25 \quad (4)$$

由式(4)得到的距离参数,从欧氏距离方面表示了标准动作和实时动作之间的差异度。另外,汉明距离为

$$D_2 = \sum_{i=1}^{25} \left( \begin{cases} 0, & d_{bi} = d_{si} \\ 1, & d_{bi} \neq d_{si} \end{cases} \right), \quad i = 1, 2, \dots, 25 \quad (5)$$

使用式(5)来评价动作特征相似度是合理的:通过比较关节点每一位是否相同,若不同则汉明距离加 1,这样得到汉明距离。关节点相似度越高,对应的汉明距离越小。它定量描述了两个动作特征之间的差异程度。

### 1.3 关节点之间的角度特征

使用余弦相似度可以从方向上区分差异,达到动作识别的目的。设通过 Kinect 得到动作关节点中相邻两个关节点位置特征点为  $A(x_i, y_i, z_i)$  和  $B(x_{i+1}, y_{i+1}, z_{i+1})$  以及对应的矢量,表示相邻两关节之间形成的向量,即

$$\mathbf{n} = (x_i, y_i, z_i) - (x_{i+1}, y_{i+1}, z_{i+1}) \quad (6)$$

对于实时获取的向量特征的夹角特征如图 2 所示。其中,  $\mathbf{n}_i$  表示两个关节点形成的第  $i$  个向量,  $\mathbf{n}_{i+1}$  表示第  $i+1$  个向量。

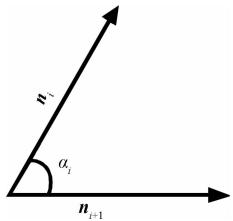


图 2 矢量之间的角度值

Fig. 2 Angle value between vectors

通过两向量之间的余弦角度计算公式(7)得到两矢量之间的角度为

$$\alpha_i = \arccos \frac{\mathbf{n}_i \cdot \mathbf{n}_{i+1}}{|\mathbf{n}_i| |\mathbf{n}_{i+1}|} \quad (7)$$

式(7)中:  $\alpha_i$  为标准动作中各个关节点组成的余弦值。

同理可以获得由实时动作形成的夹角  $\beta_i$ , 用  $\theta_i$  来表示它们之间的角度差为

$$\theta_i = |\alpha_i - \beta_i|, \quad i = 1, 2, \dots, 25 \quad (8)$$

通过式(8)获取 25 个相对角度描述符  $\mathbf{P}_2, \mathbf{P}_2 = \{\theta_1, \theta_2, \dots, \theta_{25}\}$ 。

在此,角度值  $D_3$  越大,则表明标准动作和实时动作的差异性越大:

$$D_3 = \sum_{i=1}^{25} \theta_i, \quad i = 1, 2, \dots, 25 \quad (9)$$

由式(8)得到角度参数,它从角度方面表示了标准动作和实时动作之间的差异度。

#### 1.4 特征融合动作识别评分方法

文献[9]使用关于矢量的特征进行匹配比较,可以达到较好的准确性。在这里引入文献[9]提出的改进人体动作评分方法,将评分方法拓展到适用于两种距离特征和角度特征的度量,使用距离特征和角度特征共同对实时动作识别评分:

$$S =$$

$$\begin{cases} f(a_{\max}) [(D_{st} - D) \frac{100 - S_{st}}{S_{st}} + S_{st}], & 0 \leq D \leq D_{st} \\ 0, & D \geq D_{st} \end{cases} \quad (10)$$

以及表示动作匹配度的惩罚因子  $f(a_{\max})$ :

$$f(a_{\max}) = \begin{cases} 1 - \frac{0.3}{M^2} a_{\max}^2, & 0 \leq a_{\max} \leq \sqrt{\frac{10}{3}} M \\ 0, & a_{\max} > \sqrt{\frac{10}{3}} M \end{cases} \quad (11)$$

式中:  $D$  为 1.2、1.3 节得到的距离差异度  $D_1, D_2$  和

角度差异度  $D_3$ 。最后将这 3 个匹配值乘上指定的权重后相加,得到总差异值为

$$S = S_1 W_1 + S_2 W_2 + S_3 W_3 \quad (12)$$

式(12)中:  $S_1, S_2, S_3$  分别为改进的人体动作评分方法中通过距离特征和角度特征得到的匹配度,可赋予不同的合适的权值  $W_1, W_2, W_3$ 。由于不同特征对动作相似度的贡献不同,因此该方法对动作各特征:距离特征 ( $S_1, S_2$ )、角度特征 ( $S_3$ ) 加不同的权值,从而使相似度计算达到最优。其中权值是针对不同的动作姿势通过加权平均法进行求解的,最终得到实时的匹配度。

## 2 算法分析与描述

### 2.1 算法流程

吴剑锋等<sup>[15]</sup>使用基于角度测量没有解决关节点重合的问题,并且使用单一的角度特征进行动作识别有一定的不稳定性,影响准确度。本文提出首先使用标准差和线性函数归一化的方法对获取的数据进行归一化处理,然后使用距离特征和角度特征相结合的方法进行匹配,既可以满足文献[9]所达到的实时性,又可提高动作匹配的准确度。整体的实验流程如图 3 所示。

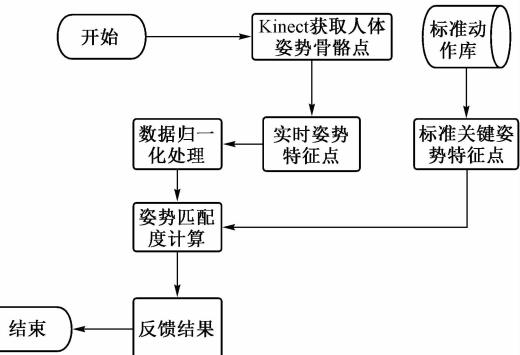


图 3 三维动作识别方法流程图

Fig. 3 Flow chart of 3D action recognition method

### 2.2 算法描述

为了验证算法的可行性,标准动作以数据库形式存在和实时动作的骨骼数据为实时获取,获取到的二者的相邻两个关节点三维坐标表示为  $(x_{ui}, y_{ui}, z_{ui}), (x_{ui}, y_{uj}, z_{uj})$  和  $(x_{ei}, y_{ei}, z_{ei}), (x_{ei}, y_{ej}, z_{ej})$ 。本文算法描述如下。

步骤 1: 使用 Kinect 提取实时动作的关节点的三维坐标。

步骤 2: 读取标准动作关节点的三维坐标。

步骤 3: 运用式(3)和式(8)获得各个距离值和角度值。距离描述符  $\mathbf{P}_1$  和角度描述符  $\mathbf{P}_2$ 。

步骤 4: 根据式(4)、式(5)和式(9)计算欧式距

离特征和汉明距离的差异度  $D_1, D_2$  以及角度特征的差异度  $D_3$ 。

步骤 5: 根据步骤 4 中得到的  $D_1, D_2$  和  $D_3$  的值, 通过评分方法式(10)、式(11)得到对应的匹配度  $S_1, S_2, S_3$ 。

步骤 6: 最后根据式(12), 两个计算公式的预设权值比重得到最终识别的动作, 获取匹配度  $S$ , 算法结束。

### 3 实验结果与分析

#### 3.1 实验设计及环境

为了体现本文算法的可行性、准确性和实时性, 分别进行了两组对比实验。对比实验是文献[9]提出的基于矢量的匹配方法和朱大勇等<sup>[11]</sup>提出的基于角度的匹配方法。

实验的硬件平台是: Kinect for Windows 摄像机一台, 深度图像的分辨率为  $512 \times 424$ , 彩色图像的分辨率为  $1920 \times 1080$ ; 实验所用计算机为戴尔 G3579, 主要配置包括: Intel(R) Core(TM) i7 8 代, 内存为 8 GB, 显卡为 NVIDIA GeForce GTX 1080ti; 系统开发环境为 Windows 10 操作系统、Visual Studio 2013 和 Kinect SDK v2.0。

#### 3.2 验证实验

使用 10 组实验来验证本文方法的有效性。经过多次实验及分析发现, 当  $S \geq 70$  时, 认为动作匹配。实验的识别效果达到最佳。并且, 当动作匹配识别准确度  $S \geq 85$  时, 则可以判定标准动作和实时动作之间实现最近匹配识别, 实时显示匹配度。

实验抽取了 30 个不同身高与体形的志愿者进行测试。每个人测试 10 种不同的动作, 匹配所定义的动作。所涉及的动作举例如图 4 所示, 图片上方显示是识别的动作结果提示, 下方显示由 Kinect 实时获取到的人物动作图和 25 个骨骼关节点连线图。图 4 为不同的实验动作识别的效果举例。

根据不同动作的定义, 在相同实验环境下与单一使用距离特征和单一使用角度特征进行对比实验。实验结果如图 5 所示。

对比文献[7]和文献[9]的实验, 选取 10 组动作与之对比, 得出以下结论。

(1) 动作 1、2、3、7、8、9、10 基于特征融合算法与单一角度测量和距离测量的识别效果识别率有一定的提升。

(2) 动作 4、5、6 从图 5 中可以看出, 相较于单一特征进行动作识别十组动作识别准确率较高, 可以达到 94% 以上的识别率。

(3) 通过实验数据可知, 文献[11]实验动作识

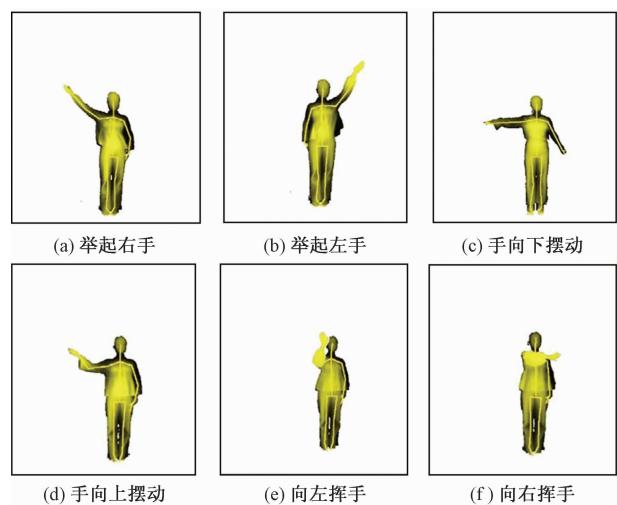


图 4 实验识别动作举例  
Fig. 4 Examples of experimental recognition actions

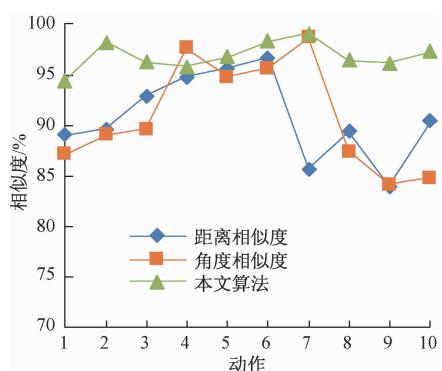


图 5 三种方法识别率对比  
Fig. 5 Comparison of recognition rate of three methods

别平均时间为 32.4 ms, 本文算法的平均识别时间为 16.7 ms。因此本文算法同时拥有较高的实时性。

### 4 结论

提出了一种基于 Kinect 的人体三维姿势实时动态识别方法。该方法使用 Kinect 对人体骨骼数据进行提取, 获取空间位置的距离特征、空间位置的角度特征, 与模板动画进行实时动态匹配, 拥有较好的识别效果。同时, 该方法算法简单, 计算速度较快, 可以满足对于姿势匹配的实时性要求, 并且使用 Unity3D 软件可自动生成动画、动态与识别人物进行匹配, 达到了动作识别的要求。但是对于 Kinect 提取动作重合的关节点处理仍然不够好, 采用 Kinect 获取人体骨骼三维特征满足了对数据获取的准确性, 但是仍然没有达到最低成本和方便性的要求。下一步工作将针对关节点重合问题, 采取新的算法或者优化方法进行处理, 对于动作识别后

与标准动作的差异问题有更好的反馈,进行动作的矫正,最终实现基于普通摄像头的人体三维动作的实时识别。

## 参 考 文 献

- 1 Cao Z, Simon T, Wei S E, et al. Realtime multi-person 2d pose estimation using part affinity fields[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 7291-7299.
- 2 Chen X, Lin K Y, Liu W, et al. Weakly-supervised discovery of geometry-aware representation for 3d human pose estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 10895-10904.
- 3 Fang H S, Xie S, Tai Y W, et al. Rmpe: regional multi-person pose estimation[C]//Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 2334-2343.
- 4 周彤彤,刘更谦,李浩源,等.基于 Kinect 的双臂机器人动作模拟与轨迹规划[J].科学与技术与工程,2017,17(7):58-62.  
Zhou Tongtong, Liu Gengqian, Li Haoyuan, et al. Dual-arm robot motion simulation and trajectory planning based on kine-ct[J]. Science Technology and Engineering, 2017, 17(7) : 58-62.
- 5 蒲兴成,王涛,张毅.基于改进 Hu 矩算法的 Kinect 手势识别[J].计算机工程,2016,42(7):165-172.  
Pu Xingcheng, Wang Tao, Zhang Yi. Kinect gesture recognition based on improved Hu moment algorithm[J]. Computer Engineering, 2016, 42(7) : 165-172.
- 6 刘杰,黄进,田丰,等.连续交互空间下的混合手势交互模型[J].软件学报,2017,28(8):2080-2095.  
Liu Jie, Huang Jin, Tian Feng, et al. Mixed gesture interaction model in continuous interaction space[J]. Journal of Software, 2017, 28 (8) : 2080-2095.
- 7 战荫伟,于芝枝,蔡俊.基于 Kinect 角度测量的姿态识别算法[J].传感器与微系统,2014,33(7):129-132.  
Zhan Yinwei, Yu Zhizhi, Cai Jun. Gesture recognition algorithm based on Kinect angle measurement[J]. Transducer and Microsystem Technologies, 2014, 33(7) : 129-132.
- 8 王怡,朱晓文,曲成璋.基于 Kinect 的健身动作识别与评价[J].计算机科学与应用,2018,8(7):1134-1145.  
Wang Yi, Zhu Xiaowen, Qu Chengzhang. Fitness movement recognition and evaluation based on Kinect[J]. Computer Science and Application, 2018, 8(7) : 1134-1145.
- 9 李红波,李双生,孙舶源.基于 Kinect 骨骼数据的人体动作姿势识别方法[J].计算机工程与设计,2016,37(4):969-975.  
Li Hongbo, Li Shuangsheng, Sun Boyuan. Posture recognition of human body based on Kinect bone data[J]. Computer Engineering and Design, 2016, 37(4) : 969-975.
- 10 辛义忠,邢志飞.基于 Kinect 的人体动作识别方法[J].计算机工程与设计,2016,37(4):1056-1061.  
Xin Yizhong, Xing Zhifei. Human motion recognition method based on Kinect[J]. Computer Engineering and Design, 2016, 37 (4) : 1056-1061.
- 11 朱大勇,郭星,吴建国.基于 Kinect 三维骨骼节点的动作识别方法[J].计算机工程与应用,2018,54(20):157-163.  
Zhu Dayong, Guo Xing, Wu Jianguo. Action recognition method based on Kinect 3D skeleton node[J]. Computer Engineering and Application, 2018, 54(20) : 157-163.
- 12 Qin C, Song A, Wu C, et al. Scenario interaction system of rehabilitation training robot based on Unity3D and Kinect[J]. Chinese Journal of Scientific Instrument, 2017, 38(3) : 530-536.
- 13 Chao L Q, Ai G S, Chang C W, et al. Scenario interaction system of rehabilitation training robot based on Unity3D and Kinect[J]. Chinese Journal of Scientific Instrument, 2017, 38(3) : 530-534.
- 14 王松,党建武,王阳萍,等.实时动作识别方法研究[J].计算机工程与应用,2017,53(3):28-31.  
Wang Song, Dang Jianwu, Wang Yangping, et al. Research on real-time action recognition method[J]. Computer Engineering and Application, 2017, 53(3) : 28-31.
- 15 吴剑锋,马梦鑫,蒋蒙婷,等.基于 Kinect 的坐姿意图判断及其应用研究[J].计算机应用与软件,2018,35(10):200-205.  
Wu Jianfeng, Ma Mengxin, Jing Mengting, et al. Study on sitting intention judgment and its application based on Kinect[J]. Computer Application and Software, 2018, 35(10) : 200-205.
- 16 吴誉兰,章小宝.基于 Kinect 的动态手臂三维姿势的识别与仿真[J].计算机仿真,2016(7):369-372.  
Wu Yulan, Zhang Xiaobao. Kinect based 3D pose recognition and simulation of dynamic arm[J]. Computer Simulation, 2016(7) : 369-372.
- 17 张全贵,蔡丰,李志强.基于耦合多隐马尔可夫模型和深度图像数据的人体动作识别[J].计算机应用,2018,38(2):454-457.  
Zhang Quangui, Cai Feng, Li Zhiqiang. Human motion recognition based on coupled multi hidden Markov model and depth image data [J]. Computer Application, 2018, 38(2) : 454-457.
- 18 谭励,杨朝玉,杨明华,等.有向移动传感器网络三维空间目标自主覆盖算法[J].计算机工程,2018,44(5):71-77.  
Tan Li, Yang Chaoyu, Yang Minghua, et al. Autonomous coverage algorithm for three-dimensional space targets in directed mobile sensor networks[J]. Computer Engineering, 2018, 44(5) : 71-77.
- 19 黄文明,莫阳.基于文本加权 KNN 算法的中文垃圾短信过滤[J].计算机工程,2017,34(3):193-199.  
Huang Wenming, Mo Yang. Chinese spam SMS filtering based on text weighted KNN algorithm[J]. Computer Engineering, 2017, 34 (3) : 193-199.
- 20 Khaire P, Kumar P, Imran J. Combining CNN streams of RGB-D and skeletal data for human activity recognition[J]. Pattern Recognition Letters, 2018, 115: 107-116.