



DOI: 10.12086/oe.2021.200099

## 基于单样本学习的多特征人体姿态模型识别研究

李国友, 李晨光\*, 王维江, 杨梦琪, 杭丙鹏

燕山大学工业计算机控制工程河北省重点实验室, 河北 秦皇岛 066004



**摘要:** 随着人机交互、虚拟现实等相关领域的发展, 人体姿态识别已经成为热门研究课题。由于人体属于非刚性模型, 具有时变性的特点, 导致识别的准确性和鲁棒性不理想。本文基于 KinectV2 体感摄像头采集的骨骼信息, 结合人体角度和距离特征, 提出了一种基于单样本学习的模型匹配方法。首先, 通过对采集的骨骼信息进行特征提取, 计算关节点向量夹角和关节点的位移并设定阈值, 其次待测姿态与模板姿态进行匹配计算, 满足阈值限定范围则识别成功。实验结果表明, 该方法能够实时的检测和识别阈值限定范围内定义的人体姿态, 提高了识别的准确性和鲁棒性。

**关键词:** 姿态模型; 骨骼数据; 单样本学习; 模型匹配; KinectV2

**中图分类号:** TP391.4; TP181

**文献标志码:** A

李国友, 李晨光, 王维江, 等. 基于单样本学习的多特征人体姿态模型识别研究[J]. 光电工程, 2021, 48(2): 200099

Li G Y, Li C G, Wang W J, et al. Research on multi-feature human pose model recognition based on one-shot learning[J].

Opto-Electron Eng, 2021, 48(2): 200099

## Research on multi-feature human pose model recognition based on one-shot learning

Li Guoyou, Li Chenguang\*, Wang Weijiang, Yang Mengqi, Hang Bingpeng

Key Laboratory of Industrial Computer Control Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China

**Abstract:** With the development of human-computer interaction, virtual reality, and other related fields, human posture recognition has become a hot research topic. Since the human body belongs to a non-rigid model and has time-varying characteristics, the accuracy and robustness of recognition are not ideal. Based on the KinectV2 somatosensory camera to collect skeletal information, this paper proposes a one-shot learning model matching method based on human body angle and distance characteristics. First, feature extraction is performed on the collected bone information, and the joint point vector angle and joint point displacement are calculated and a threshold is set. Secondly, the pose to be measured is matched with the template pose, and the recognition is successful if the threshold limit is met. Experimental results show that the method can detect and recognize human poses within the defined threshold in real-time, which improves the accuracy and robustness of recognition.

**Keywords:** pose model; skeleton data; one-shot learning; model matching; KinectV2

收稿日期: 2020-03-23; 收到修改稿日期: 2020-06-02

基金项目: 河北省高等学校科学技术研究青年基金项目(2011139); 河北省自然科学基金项目(F2012203111)

作者简介: 李国友(1972-), 男, 博士, 教授, 主要从事机器视觉和图像处理算法的设计与研究。E-mail: lgyysu@163.com

通信作者: 李晨光(1994-), 男, 硕士研究生, 主要从事图像处理和模式识别的研究。E-mail: 1498578260@qq.com

版权所有©2021 中国科学院光电技术研究所

## 1 引言

人体姿态识别是计算机视觉研究的热门课题,为人机交互、虚拟现实等领域的发展提供技术支持<sup>[1]</sup>。人体运动动作的时变性以及运动场景的复杂变化性,导致人体姿态识别的实时性、准确性和鲁棒性不理想。因此,如何实时准确地捕捉和识别人体姿态在计算机视觉领域引起了极高的关注。

目前,人体姿态识别研究方法主要有基于统计的算法<sup>[2]</sup>(隐马尔科夫模型(HMM)、动态贝叶斯网络(DBN)),基于模板的算法(模型、动态规划、动态时间规整<sup>[3]</sup>(DTW)),以及基于语法的算法(有限状态机(FSM)、上下文无关文(CFG))。在针对人体姿态研究的算法中,基于模板匹配相比其他算法具有更好的识别准确性和鲁棒性,但是模型的建立需要大量数据且特征计算复杂<sup>[4]</sup>。李红波等<sup>[5]</sup>提出了基于 Kinect 骨骼数据的静态三维手势识别,通过获取手部角度特征向量确定待测姿态模板,待测姿态与姿态模板进行最佳姿态匹配,由于行为时间间隔的持续性和图像噪声影响,每帧图像无法实时地检测识别。Zhe 等人<sup>[6]</sup>提出了零件相似性的实时多人人体姿态识别,该算法采用自下而上(down-top approach)方法,首先对人体关键点进行检测,其次建立矢量场推测关键部位并进行连接,最后采用深度学习网络结构构建特征图像集进行姿态预测;但该算法存在一定的局限性,只考虑了肢体关键点的位置信息,忽略了旋转、尺度不变性,因而只能在局部区域进行姿态预测连接。Pfischer 等人<sup>[7]</sup>提出了基于卷积神经网络和 FastDTW 的 Kinect 手势控制机器人的移动,对从 Kinect 上捕捉获取图像的每帧关节信息进行提取和处理,建立姿态动作所有帧派生图像,采用联合训练和个人训练进行卷积神经网络(CNN)训练,最后将移动机器人系统添加到训练好的系统中,根据姿态算法的分类结果控制机器人的移动。赵海勇等人<sup>[8]</sup>提出了基于视频流的运动人体行为识别,采用帧间差分法和改进 C-V 模型算法进行运动人体分割,检测运动目标的轮廓曲线,对分割的轮廓进行加权运动串分类,构建动作串时对不同的姿态赋予不同的权重,最后通过对全局运动信息和局部特征信息的异常行为进行识别,该方法可有效地将全局信息和局部特征信息进行互补加权串联融合,特征点更加明显,识别效果更好。Xu 等人<sup>[9]</sup>提出了多级深度运动图的人体动作识别,使用多级帧选择采样模型(MFSS),首先将输入的深度图像投影到三个正交的笛卡尔平面上,其

次采用动态图和静态图(MSM)映射方法获得动态历史图像和静态历史图像来表示姿态,利用 LBP 提取块的 MSM 的特征信息和使用 Fisher 内核整合块,最后采用极限学习机(KELM)分类器进行姿态识别。Vinyals 等<sup>[10]</sup>提出了小样本学习匹配网络,设计了匹配网络(matching networks,新的神经网络结构),该网络将标记的实例映射到已标记的小型支持集,且标记的样本无需调整就能适应新的类型,并且损失函数使预测值和真实值的误差最小化。现阶段人体姿态识别的研究取得了不错的成果,但基于模板匹配的研究忽略了姿态的时序性,对于姿态的瞬变性仍然存在技术漏洞。

现阶段对人体姿态识别仍然存在未定义姿态干扰等技术缺陷,导致识别的实时性和准确性在一定程度上没有达到预期效果。因此,本文针对模型特征提取和模型数据集训练,基于单样本学习对获取的图像进行映射训练,采用关节向量角度和距离特征以及关节成角平面与 Kinect 的 XOZ 平面夹角,保证姿态特征的不变性,同时在姿态识别过程中保证特征点的高契合度。该方法对定义好的姿态模型识别准确率达到 98.29%,实时性也有所提高。

## 2 人体骨骼数据集

### 2.1 人体关节简介

KinectV2 获取深度图像的原理是采用飞行时间(TOF)技术<sup>[11]</sup>,传感器发射红外光通过物体反射的时间和相位差来确定深度信息。尤其在骨骼检测和跟踪方面有着强大的功能,可捕捉到 25 个关节信息,这 25 个关节能实时有效地反映出人体任意姿态<sup>[12]</sup>,25 个骨骼关节如图 1 所示。

图中蓝色标注代表姿态的角度和距离特征的相关关节,这些关节依次为 SpineBase(脊柱基部)、HandLeft(左手)、ElbowLeft(左肘)、ShoulderLeft(左肩)、ShoulderRight(右肩)、ElbowRight(右肘)、HandRight(右手)。姿态的特征计算基于标注关节的三维坐标实时信息。

### 2.2 人体动作识别数据集

KinectV2 动作数据识别数据集<sup>[13]</sup>可分为五种类型,每种类型既有相通性又有区别性,GAMING DATASRTS-G3D 数据集定义规划了 20 个游戏动作;UTKINECT-ACTION3D DATASET 数据集基于 RGB、深度和骨架节点位置定义了 10 种动作;FLORENCE 3D ACTIONS DATASET 数据集定义了 9 个动作;MSR

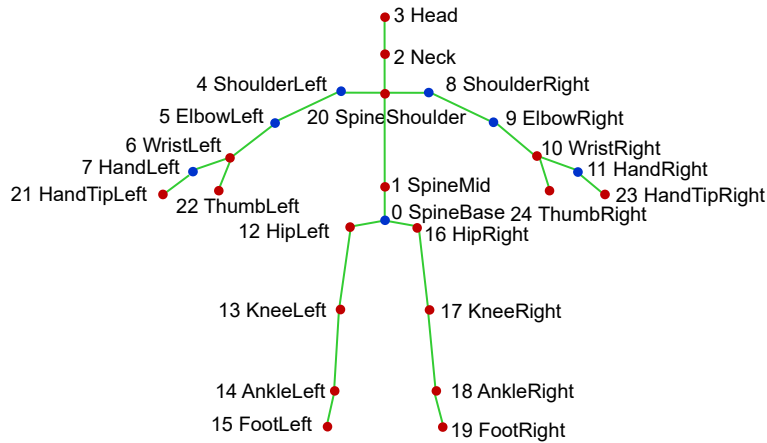


图 1 人体骨骼关节示意图

Fig. 1 Schematic diagram of human bone joints

ACTION3D 数据集通过记录人体动作序列,按动作的时序性记录了 20 个动作; NTU RGB+D ACTION RECOGNITION DATASET 数据集利用 RGB 视频、深度图序列、3D 骨骼数据和红外视频捕捉动作,但是该数据集依赖于三个 KinectV2 摄像机同时捕捉,捕捉难度和数据处理分析难度大。动作类型如图 2 所示。

### 3 骨骼信息的多特征人体姿态模型识别

#### 3.1 基本流程

人体姿态识别的核心内容为: 首先,进行姿态捕

捉、特征提取并标定,其中特征标定是将姿态动作标记为一个静态标签的过程,一个静态标签对应一个特定的姿态。其次,采用单样本学习模型匹配算法将待识别特征样本与训练完成的特征模板库进行角度和距离特征匹配计算,并根据设定的阈值范围进行契合度对比,阈值范围通过多次实验确定。最后,显示识别结果的过程。图 3 为人体姿态识别流程图。

#### 3.2 人体姿态特征提取

由于骨骼信息的时变性,如何将不稳定的骨骼信息进行特征标定并转换得到构建模型的特征信息是模型识别的关键。为使获取的特征信息具有良好的鲁棒

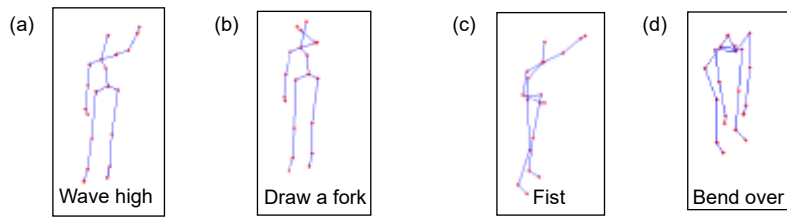


图 2 动作类型示意图

Fig. 2 Schematic diagram of action types

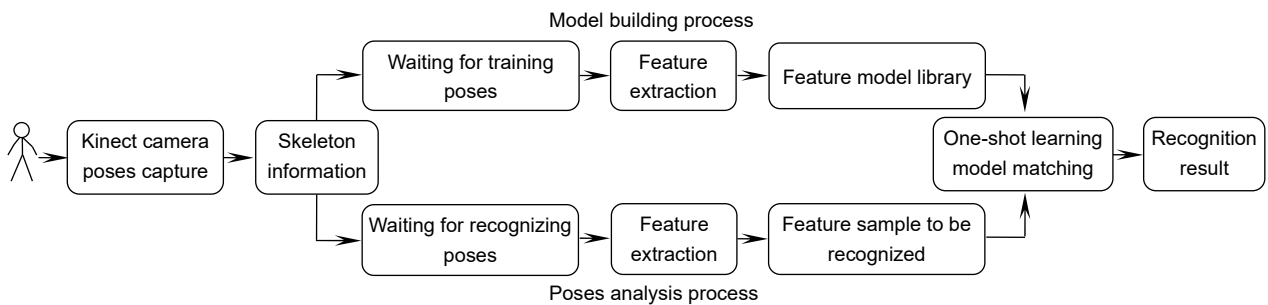


图 3 人体姿态识别流程图

Fig. 3 Flow chart of gesture recognition

性, 本文提出了角度特征和距离特征。

### 3.2.1 角度特征

由于人体运动具有高自由度和非刚性的特点, 导致人体骨骼关节运动随意性很大, 没有特定的姿态作为匹配对象很难去识别, 因此选取骨骼信息作为技术支撑<sup>[14]</sup>。在 KinectV2 骨骼坐标系下的 25 个关节点都有独立的三维坐标信息, 若直接采用 25 个关节点的三维信息构建人体姿态, 数据集维度达到 75 维, 姿态表征信息冗余, 计算难度非常大。过多的关节点表征使识别结果容易错乱而达不到预期效果。角度特征的提取是基于骨骼结构的不变性, 关节点向量之间可以随机组合成不同的向量角, 向量角既可以高度表征姿态信息, 又可以将与姿态动作无关的信息进行剔除, 减少了数据集维度, 可有效利用骨骼信息, 从而使识别准确性更高。

本文选取角度组合数量多的上肢进行特征提取, 基于关节点的三维坐标值构建关节点向量, 然后对关节向量组合成的夹角进行计算。本文选取 HandLeft-ElbowLeft-ShoulderLeft 构建左上肢向量角和 ShoulderRight-ElbowRight-HandRight 构建右上肢向量角, 对两个关节向量角分别设定不同的角度构建姿态特征模板库。关节向量角如图 4 所示,  $\theta_1$  为构建的左上肢向量角,  $\theta_2$  为右上肢向量角。

图 5 为构建的左上肢向量角示意图, 计算过程如下:

- 1) 获取关节点三维坐标信息如下:

HandLeft  $(a_1, b_1, c_1)$ , ElbowLeft  $(a_2, b_2, c_2)$ , ShoulderLeft  $(a_3, b_3, c_3)$ ;

- 2) 构建关节点向量, 定义左手肘到左手关节向量及左手肘到左肩膀关节向量如下:

$$\alpha = (a_1 - a_2, b_1 - b_2, c_1 - c_2), \quad (1)$$

$$\beta = (a_3 - a_2, b_3 - b_2, c_3 - c_2) \quad (2)$$

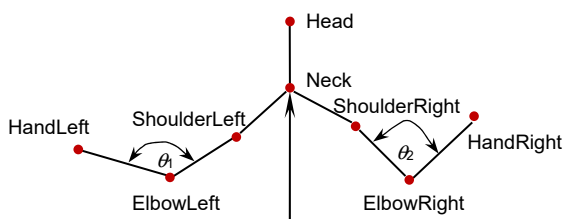


图 4 关节向量角示意图

Fig. 4 Schematic diagram of joint vector angle

- 3) 计算向量  $\alpha$  和  $\beta$  组成的向量夹角  $\theta$  值:

$$\theta = \arccos \frac{|\alpha \cdot \beta|}{\|\alpha\| \times \|\beta\|}, \quad (3)$$

其中  $\alpha \cdot \beta$  为向量内积,

$$\|\alpha\| = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2 + (c_1 - c_2)^2}, \quad (4)$$

$$\|\beta\| = \sqrt{(a_3 - a_2)^2 + (b_3 - b_2)^2 + (c_3 - c_2)^2} \quad (5)$$

基于式(1)~式(5)的计算提取角度特征构建人体姿态数据集<sup>[15]</sup>, 在姿态匹配识别过程, 设定阈值  $d$ , 当  $\theta$  计算值满足取值范围  $(\theta - d) \leq \theta \leq (\theta + d)$ , 即识别成功。

### 3.2.2 距离特征

角度特征可以有效地描述人体姿态类型, 但在某些情况, 采用距离特征能简单快速地描述出姿态, 而且能准确标记出人体关节点的相对位置, 避开了角度特征的干扰性。基于关节点特征的提取, 本文选取了人体骨骼比较稳定的躯干和上肢关节点, 也是识别效果最佳的关节点。

本文选取 SpineBase(脊柱基部)作为参考点, Neck(脖颈)作为平面映射点, HandLeft(左手)和 HandRight(右手)为距离点进行距离标定, 然后计算距离点到参考点三维坐标距离。由于距离特征维度上的差异性和分离性, 采用了 X 方向上的距离作为特征, 距离特征如图 6 所示。

特征距离计算过程如下:

- 1) 获取关节点的三维坐标信息如下:

SpineBase  $(x_1, y_1, z_1)$ , Neck  $(x_2, y_2, z_2)$ , HandLeft  $(x_3, y_3, z_3)$ , HandRight  $(x_4, y_4, z_4)$ ;

- 2) 计算 X 方向上的距离, 选取脖颈为脊柱基点映射点, X 方向上的坐标一致, 只需计算手部位置到脖颈的距离。即距离特征  $x_1 = x_2$ , 计算如下:

左手到脊柱基点的距离:

$$L_1 = |x_3 - x_2| \quad (6)$$

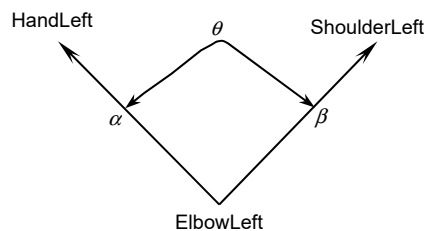


图 5 左上肢向量角

Fig. 5 Vector angle of left upper limb

右手到脊柱基点的距离:

$$L_2 = |x_4 - x_2| \quad (7)$$

3) 设定定值  $f_1$  和  $f_2$ , 比较  $(L_1 > f_1) \cup (L_2 < f_2)$  或  $(L_1 < f_1) \cup (L_2 > f_2)$ , 若满足设定条件即特征对比成功。本文采用综合角度特征和距离特征结合的方法, 对人体姿态进行准确的标定和描述, 具有旋转缩放不变性, 满足了不同人、不同角度和不同位置的姿态识别。

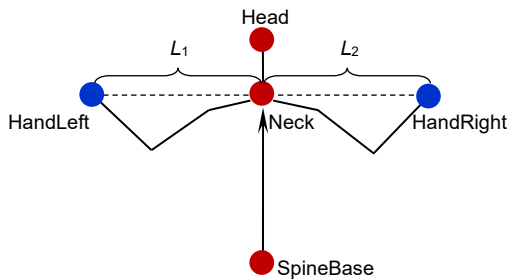


图 6 距离特征示意图

Fig. 6 Schematic diagram of distance characteristics

### 3.3 姿态识别算法

在人体姿态识别过程中, 姿态特征提取和动作识别尤为重要。本文提出了基于人体姿态关节角度和距离特征的单样本学习的模型匹配算法进行动作识别。该算法的核心思想是监督学习领域的单样本学习归类, 对小样本进行训练归类, 即训练一个数据端到另一个数据端的分类器<sup>[16]</sup>, 训练的样本库模型具有高度不变性, 模型匹配方法将训练完成的数据库模板与识别过程的静态数据做欧氏距离计算, 选择契合度最高的样本模板所属类型作为识别结果。该算法具有计算复杂度低、实时性和准确性高的特点。

单样本学习作为小样本学习的特殊情况, 每种类型只需提供一个样本作为参考模板无需依赖海量的样本数据。在单样本学习的训练集中设定了多种类型, 每种类型对应一个样本, 在训练过程中, 随机抽取训练集中的  $C$  种类型, 每种类型一个(one)样本(所有的特征点集合)构建 meta-task<sup>[17]</sup>, 作为模型的 support set(支持集)<sup>[18]</sup>输入; 然后从设定的类型数据抽取 batch(批)样本作为模板的 batch set<sup>[19]</sup>(预测对象), 这种训练模式称为 one-way one-shot 单样本学习。经过多次训练, 可得到包含一个类型所有特征点的 meta-task。单样本学习模型结构如图 7 所示。

对建立好的模型结构进行参数转换, 寻找已训练数据集和将要预测实验数据集的相似度, 算法流程如下:

1) 训练过程: 定义一个包含多个样本数据的支持集:

$$S = \{(x_i, y_i)\}_{i=1}^{|k|} \quad (8)$$

将测试样本数据  $\hat{x}$  进行分类进而得到他们各自的标签  $\hat{y}$ , 即:

$$\hat{y} = \sum_{i=1}^{|k|} a(\hat{x}, x_i) y_i \quad (9)$$

其中:  $a$  用来计算度量测试样本和训练样本的契合度, 即:

$$a(\hat{x}, x_i; \theta) = \frac{\exp(c(f(\hat{x}), g(x_i)))}{\sum_{i=1}^{|k|} \exp(c(f(\hat{x}), g(x_i)))} \quad (10)$$

其中:  $f$  为模型结构中测试样本编码,  $g$  为训练样本编码,  $c$  为计算测试样本到训练样本的匹配距离<sup>[22]</sup>。

定义一个支持集到测试数据映射为

$$P_\theta(\hat{y} | \hat{x}, S) = \sum_{(x_i, y_i) \in S} k_\theta(x, x_i, S) y_i \quad (11)$$

输入一个测试数据, 计算测试数据到支持集  $S$  的距离度量<sup>[23]</sup>:

$$y = P(\cdot | \hat{x}, S) = \sum_{i=1}^{|k|} a(\hat{x}, x_i; \theta) y_i \quad (12)$$

将支持集样本数据嵌入模型  $g$  反复优化, 也可通过修改测试样本嵌入模型  $f$  将提取的特征反复优化后得到每个姿态在不同情况下的数据集, 即训练过程结束。

2) 编码样本集: 对于训练样本数据集映射到预测数据的序列问题, 本文选取长短期记忆(LSTM), 旨在解决训练样本冗余数据的干扰、匹配过程中动作间隔和延迟时间的问题<sup>[25]</sup>。模型结构中  $g(\theta)$  是一个双向 LSTM, 作为单输入(对应支持集  $S$  中的每个样本  $(x_0, x_1, x_2, \dots)$  双输出  $(m_i, n_i)$  的典型递归神经网络, 将  $x_i$  输入到 LSTM 进行第一个编码记为  $g'(x_i)$ , 然后对于支持集  $S$  中的样本  $x_i$  进行递归编码:

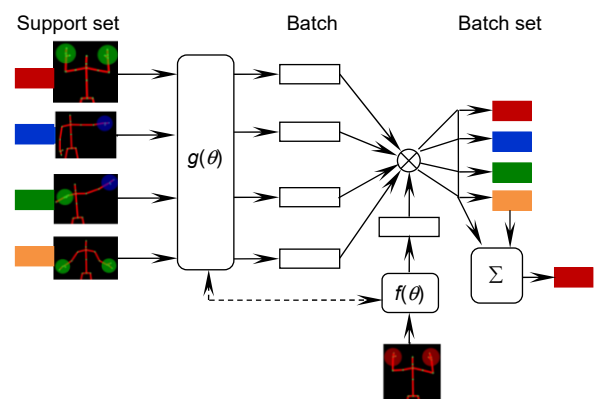


图 7 单样本学习模型结构

Fig. 7 One-shot learning model structure

$$g(x_i, S) = \bar{m}_i + \bar{n}_i + g'(x_i) \quad (13)$$

$$\bar{m}_i, \bar{n}_i = M_L(f'(\hat{x}), \bar{m}_{i-1}, \bar{n}_{i-1}) \quad (14)$$

$$\bar{m}_i, \bar{n}_i = M_L(g'(x_i), \bar{m}_{i+1}, \bar{n}_{i+1}) \quad (15)$$

其中:  $M_L$  表示长短期记忆函数,  $m_i, n_i$  为双向编码支持集。

3) 编码测试集: 考虑到支持集  $S$  中样本  $x_i$  的随机性, 可能在一定程度上出现序列紊乱现象, 因此设定一个嵌入函数  $f(\theta)$ :

$$f(\hat{x}, S) = M_L(f'(\hat{x}), g(S), K) \quad (16)$$

$f'(x)$  为嵌入层的输出,  $K$  为迭代步数。

$$\hat{m}_k, n_k = M_L(f'(\hat{x}), m_{k-1}, n_{k-1}) \quad (17)$$

$$m_k = \hat{m}_k + f'(\hat{x}) \quad (18)$$

在迭代的每一步中, 嵌入层  $f'(x)$  输出值不变, 保证了测试数据的完整性。

## 4 实验验证和分析

### 4.1 实验环境

本次实验硬件基于 Windows 10 企业版的 64 位操作系统, 处理器为 Intel(R) Core(TM) i5-6400 CPU 主频 2.71 GHz, 搭配内存为 8 GB, 显卡为 NVIDIA

GeForce GTX 750 Ti, 二代 Kinect for Windows 体感摄像头; 软件基于 Visual Studio 2017, Kinect SDK2.0, OpenCV3.4.1 视觉库。

### 4.2 实验设计

本文通过对人体主要骨骼关节点进行特征提取, 建立模板库, 采用单样本学习的模型匹配算法, 实现人体姿态识别。基于角度和距离特征定义了 7 种姿态, 通过姿态识别结果验证本文算法的准确性、实时性和鲁棒性。

### 4.3 姿态识别实验

本次实验对定义的姿态进行识别验证, 为了使识别效果更好, 经过多次实验校正, 角度特征阈值  $d=20$ , 距离特征高度  $f_1=0.45$ ,  $f_2=0.2$  时, 偏差在设定值范围之内, 识别效果最佳。角度特征姿态实验结果如图 8 所示, 图中左侧字母代表姿态标记, 圆框代表模板库样本姿态, 骨骼和彩色图像代表实时测试姿态, 方框显示动作名称。

距离特征姿态识别实验结果如图 9 所示, 图片显示内容和角度特征一致, 距离特征只需计算左右手关

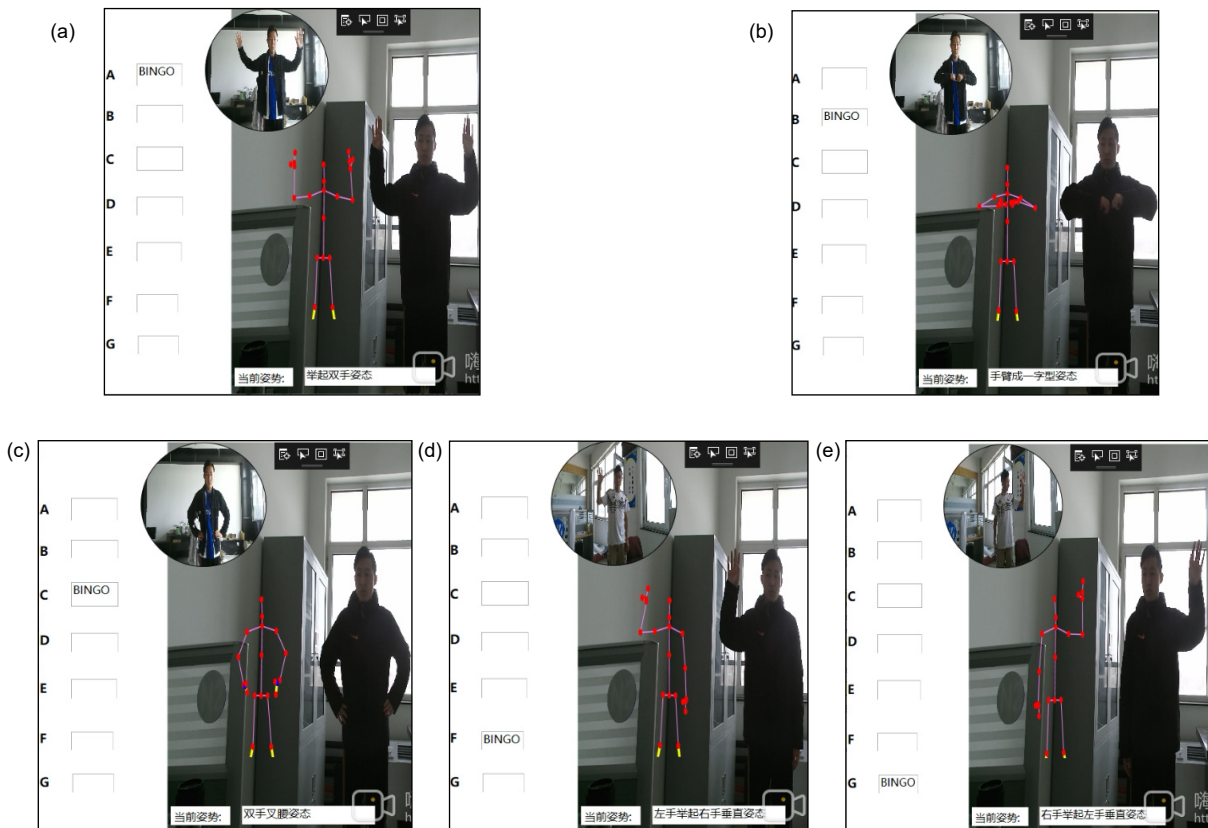


图 8 角度特征姿态识别结果示意图

Fig. 8 Schematic diagram of recognition results of angular features

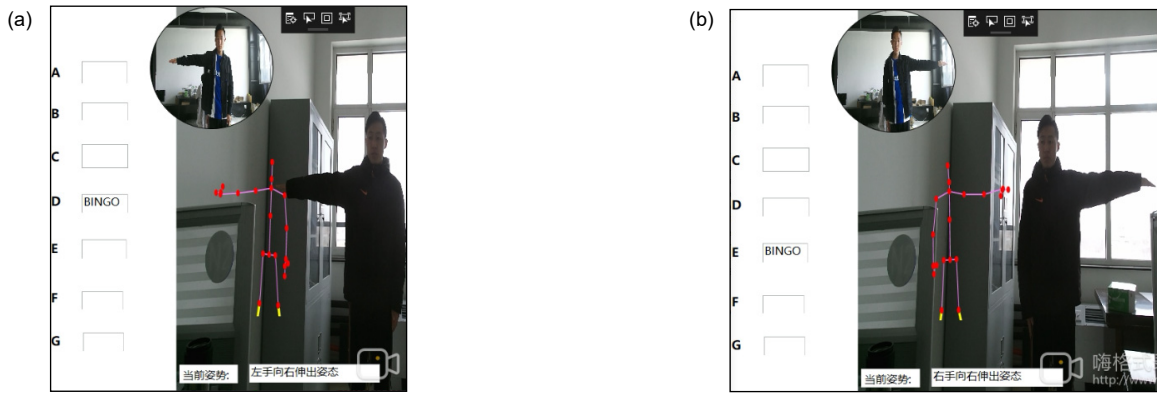


图 9 距离特征姿态识别结果示意图

Fig. 9 Schematic diagram of distance feature pose recognition results

节点到脊柱基部的设定距离，即可完成姿态识别。

#### 4.4 实验结果分析

本文定义的姿态动作标记为 A(举起双手)、B(手臂成一字型在胸前)、C(双手叉腰)、D(左手向左伸直)、E(右手向右伸直)、F(左手举起成 90°右手自然下垂)、G(右手举起成 90°左手自然下垂)7 种。为验证本文算法的准确性、实时性和鲁棒性，实验过程中 5 人进行不同位置、不同身高体型、不同角度的实验，姿态动作 10 次/人，实验结果如表 1 所示。

从表 1 中可看出，动作 F 和 G 识别错误率相对较高。通过分析可知原因有两点：1) 特征点少，存在匹配误差；2) 测试者没有做出识别范围内的动作。实验结果表明，本文基于角度特征和距离特征的平均识别率高达 98.29%，准确率较高。

#### 4.5 算法对比

针对算法的识别率和实时性，选取六种算法作为对比对象，对比算法如下：

1) 算法 1：文献[26]提出的通过提取梯度局部自

相关(GLAC)特征来表征动作相对应的运动历史图像(MHI)和静态历史图像(SHI)联合表征动作的算法。

2) 算法 2：文献[7]提出的基于 CNN 和 FastDTW 进行联合训练和个别训练的方法。

3) 算法 3：文献[21]提出的姿态图像生成特征代表特定的身体部位，然后将生成的特征馈送到 SVM 中生成姿态学习，识别预定义姿态模型的算法。

4) 算法 4：文献[20]提出的基于分割技术将动态表示和骨架特征序列匹配的识别方法。

5) 算法 5：文献[24]提出的使用 Gabor 小波提取图像纹理特征和梯度空间边缘分布(SEDG)计算形状特征构建扩展多分辨特征(EMRF)模型的方法。

6) 算法 6：文献[27]提出的定义一个由三元素组成的姿态描述符，使用 Fisher Vector(FV)对描述符进行编码的算法。

为证明本文提出算法的优越性，本文提出的算法与六种算法的平均识别率和平均识别时间进行对比，对比结果如表 2 所示。

从表 2 对比可得出：

表 1 七种姿态识别结果

Table 1 Recognition results of seven gestures

Poses	Experiments	Recognition error	Recognition rate/%
A	50	0	100.00
B	50	0	100.00
C	50	1	98.00
D	50	0	100.00
E	50	1	98.00
F	50	2	96.00
G	50	2	96.00

表 2 七种算法识别率对比

Table 2 Comparison of recognition rates of seven algorithms

Algorithm type	Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4	Algorithm 5	Algorithm 6	Text algorithm
Average recognition rate/%	97.30	90.78	98.14	97.80	96.16	93.07	98.29
Average recognition time/ms	42	58	47	34	44	52	32

1) 算法 1 的平均识别率为 97.3%，该算法通过联合使用运动历史图像(MHI)和静态历史图像(SHI)两种功能的动作表示方法，首先对提取的动作进行 12-正规化协作表示分类器(12-CRC)，然后对测试动作评估识别。该算法主要提取图像的纹理特征，将特征传递给 GLAC 构建数据集，但该算法容易混淆特征相似的动作，且数据集的构建需要大量的数据作为支撑。

2) 算法 2 的平均识别率为 90.78%，该算法利用姿态的所有帧创建图像，进而卷积神经网络训练，在匹配控制中运用 FastDTW 进行识别。虽采用 CNN 对姿态训练，有效地解决了训练过程特征点丢失和冗余，但实时性和准确性并没达到预期效果。

3) 算法 3 的平均识别率逼近本文识别率，为 98.14%。该算法采用了主成分分析对姿态特征评估，将评估的特征反馈到 SVM 生成姿态学习。该算法的高识别率依据了主要关节的提取和评估，但同一主关节不同姿态存在干扰性，出现误识情况。

4) 算法 4 的平均识别率为 97.80%，该算法将骨架特征序列划分为关键姿态和运动段姿态，构建高置信度的框架集，最后利用形状动态时间规整(shapeDTW)测量两个特征序列的运动特征段距离。该算法有效地分割骨架特征，高置信度的框架集提高了识别率，但骨架特征信息具有连续性和瞬变性，影响运动特征段的提取。

5) 算法 5 的平均识别率为 96.16%，该算法通过模糊推理模型从视频序列中选取单个关键图像，然后计算选取图像的纹理和形状特征，构建多分辨率的特征描述模型并通过 SVM 分类器进行识别的算法。该算法对特征进行分类提取，且构建的多分辨率特征模型削弱了场景的干扰，但处理图像旋转缩放性能力和实时性较差。

6) 算法 6 的平均识别率为 93.07%，该算法使用 3D 骨架信息构建三元素姿态描述符，使用极限学习机(ELM)对编码的描述符进行训练。该算法采用时空动作描述符构建人体姿态框架，解决了姿态识别过程中

序列丢帧问题，但编码描述符的计算复杂，三元素的特征提取一定程度上导致识别时间增加，实时效果减弱。

本文算法的平均识别率相对高了几个百分点，针对算法 1 特征点相似，本文从角度和距离特征进行识别，拓展了姿态动作集，且对特征相似的姿态进行标定区分；有效的特征点提取解决了算法 2 数据丢失和冗余现象；解决了算法 3 中姿态互扰问题；解决了算法 4 关键特征提取问题，且本算法的容错率更高；针对算法 5 受场景影响，本算法排除了场景的干扰且具有旋转缩放不变性，解决了算法 6 中对框架模型依赖性高，且本算法特征计算量小。通过对比平均识别时间，本文算法识别时间只有 32 ms，识别时间更短。本算法不仅提高了识别率，且实时性更好。

## 5 结 论

本文基于 KinectV2 采集骨骼数据，通过对获取骨骼关节的三维坐标信息进行角度和距离特征提取，采用单样本学习的模型匹配方法，解决了特征提取和特征向量构建的问题，运用数学矢量角和余弦值将构建的特征向量组合成角度，每个姿态特征向量角设定一个匹配阈值，匹配过程中测试姿态特征向量角在阈值范围之内则匹配成功。实验结果表明，本文算法提高了人体姿态识别的抗干扰性、准确性、实时性和鲁棒性。

## 参考文献

- [1] Henry P, Krainin M, Herbst E, et al. RGB-D mapping: using Kinect-style depth cameras for dense 3D modeling of indoor environments[J]. *Int J Rob Res*, 2012, **31**(5): 647–663.
- [2] Huang G F, Li Y. A survey of human action and pose recognition[J]. *Comput Knowl Technol*, 2013, **12**(1): 133–135.  
黄国范, 李亚. 人体动作姿态识别综述[J]. *电脑知识与技术*, 2013, **12**(1): 133–135.
- [3] Shotton J, Sharp T, Kipman A, et al. Real-time human pose recognition in parts from single depth images[J]. *Commun ACM*, 2013, **56**(1): 116–124.
- [4] Yan L M, Du B, Guo Q, et al. Recognize the tilt fingertips by partial scan algorithm[J]. *Opto-Electron Eng*, 2016, **43**(12):

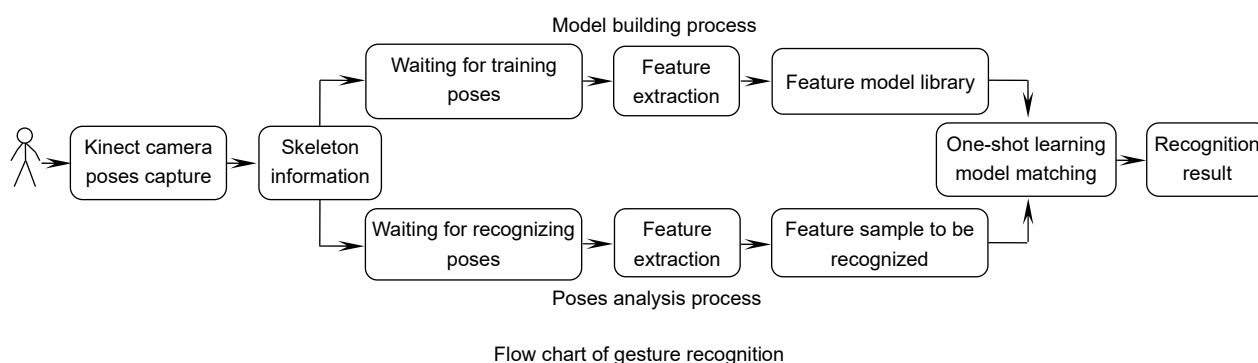


- 147–153.  
严利民, 杜斌, 郭强, 等. 基于局部扫描法对倾斜指势的识别[J]. 光电工程, 2016, **43**(12): 147–153.
- [5] Li H P, Ding L J, Wu Y, *et al.* Static three-dimensional gesture recognition based on Kinect skeleton data[J]. *Comput Appl Softw*, 2015, **14**(9): 161–165.  
李红波, 丁林建, 吴渝, 等. 基于 Kinect 骨骼数据的静态三维手势识别[J]. 计算机应用与软件, 2015, **14**(9): 161–165.
- [6] Cao Z, Simon T, Wei S E, *et al.* Realtime multi-person 2D pose estimation using part affinity fields[C]//*Conference on Computer Vision and Pattern Recognition(CVPR)*, 2017: 7291–7299.
- [7] Pfitscher M, Welfer D, Evaristo José Do Nascimento, *et al.* Article Users Activity Gesture Recognition on Kinect Sensor Using Convolutional Neural Networks and FastDTW for Controlling Movements of a Mobile Robot[J]. *Inteligencia Artificial Revista Iberoamericana de Inteligencia Artificial*, 2019, **22**(63): 121–134.
- [8] Zhao H Y. Research of human action recognition based on video stream[D]. Xi'an: Xidian University, 2011.  
赵海勇. 基于视频流的运动人体行为识别研究[D]. 西安: 西安电子科技大学, 2011.
- [9] Xu W Y, Wu M Q, Zhao M, *et al.* Human action recognition using multilevel depth motion maps[J]. *IEEE Access*, 2019, **7**: 41811–41822.
- [10] Vinyals O, Blundell C, Lillicrap T, *et al.* Matching networks for one shot learning[C]//*Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016: 3637–3645.
- [11] Zhu D Y, Guo X, Wu J G. Action recognition method using Kinect 3D skeleton data[J]. *Comput Eng Appl*, 2018, **54**(20): 152–158.  
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): 152–158.
- [12] Cai X Q, Tu Y X, Yu Y J, *et al.* Human posture recognition method based on few key frames sequence[J]. *J Graph*, 2019, **40**(3): 532–538.  
蔡兴泉, 涂宇欣, 余雨婕, 等. 基于少量关键序列帧的人体姿态识别方法[J]. 图学学报, 2019, **40**(3): 532–538.
- [13] Qian Y Z, Shen Y F. Hybrid of pose feature and depth feature for action recognition in static image[J]. *Acta Autom Sin*, 2019, **45**(3): 626–636.  
钱银中, 沈一帆. 姿态特征与深度特征在图像动作识别中的混合应用[J]. 自动化学报, 2019, **45**(3): 626–636.
- [14] Monir S, Rubya S, Ferdous H S. Rotation and scale invariant posture recognition using Microsoft Kinect skeletal tracking feature[C]//2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), 2012: 829–842.
- [15] Guo T H, Chen Y J, Lin L. Gesture recognition based on the gesture angle of dual Kinect[J]. *Sci Technol Eng*, 2019, **19**(29): 172–178.  
郭同欢, 陈姚节, 林玲. 基于姿态角的双 Kinect 数据融合技术及应用[J]. 科学技术与工程, 2019, **19**(29): 172–178.
- [16] Wang P, Liu L Q, Shen C H, *et al.* Multi-attention network for one shot learning[C]//2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 2721–2729.
- [17] Yang Y, Saleemi I, Shah M. Discovering motion primitives for unsupervised grouping and one-shot learning of human actions, gestures, and expressions[J]. *IEEE Trans Pattern Anal Mach Intell*, 2013, **35**(7): 1635–1648.
- [18] Sundermeyer M, Alkhoul T, Wuebker J, *et al.* Translation modeling with bidirectional recurrent neural networks[C]//*Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014: 14–25.
- [19] Cheng X L, He M Y, Duan W J. Machine vision based physical fitness measurement with human posture recognition and skeletal data smoothing[C]//2017 *International Conference on Orange Technologies (ICOT)*, 2017: 323–342.
- [20] Li Q M, Lin W X, Li J. Human activity recognition using dynamic representation and matching of skeleton feature sequences from RGB-D images[J]. *Signal Process Image Commun*, 2018, **68**: 265–272.
- [21] Zhang Z Q, Liu Y N, Li A, *et al.* A novel method for user-defined human posture recognition using Kinect[C]//2014 *7th International Congress on Image and Signal Processing*, 2014: 724–739.
- [22] Sagayam K M, Hemanth D J. Hand posture and gesture recognition techniques for virtual reality applications: a survey[J]. *Virtual Real*, 2017, **21**(2): 91–107.
- [23] Stephenson R M, Chai R, Eager D. Isometric finger pose recognition with sparse channel SpatioTemporal EMG imaging[J]. *Annu Int Conf IEEE Eng Med Biol Soc*, 2018, **2018**: 5232–5235.
- [24] Vishwakarma D K, Singh T. A visual cognizance based multi-resolution descriptor for human action recognition using key pose[J]. *AEU-Int J Electron Commun*, 2019, **107**: 157–169.
- [25] Liu X X, Feng X Y, Pan S J, *et al.* Skeleton tracking based on Kinect camera and the application in virtual reality system[C]//*Proceedings of the 4th International Conference on Virtual Reality*, 2018: 21–25.
- [26] Bulbul M F, Islam S, Ail H. 3D human action analysis and recognition through GLAC descriptor on 2D motion and static posture images[J]. *Multimed Tools Appl*, 2019, **78**(15): 21085–21111.
- [27] Agahian S, Negin F, Köse C. An efficient human action recognition framework with pose-based spatiotemporal features[J]. *Eng Sci Technol*, 2020, **23**(1): 196–203.

# Research on multi-feature human pose model recognition based on one-shot learning

Li Guoyou, Li Chenguang\*, Wang Weijiang, Yang Mengqi, Hang Bingpeng

Key Laboratory of Industrial Computer Control Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China



**Overview:** With the development of human-computer interaction, virtual reality, and other related fields, human pose recognition has become a hot research topic. Since the human body belongs to a non-rigid model and has time-varying characteristics, the accuracy and robustness of recognition are not ideal. Based on the KinectV2 somatosensory camera to collect skeletal information, this paper proposes a one-shot learning model matching method based on human body angle and distance characteristics. KinectV2 somatosensory camera can capture color images, depth images, and skeletal images. This article combines color images and skeletal images to extract and recognize human poses, thereby making the recognition more accurate, real-time, and robust. Through the feature extraction of the obtained bone information, the feature vector is constructed using the vector in mathematical thought, that is, the endpoint of the joint point minus the joint point of the starting point. The joint angle in this article uses the elbow joint as the fulcrum, shoulder joint, and the hand joint is constructed as the end-point. In this way, multiple types of poses can be combined, and the attitude sample library can be expanded. The joint point vector angle and the joint point displacement are calculated, and the threshold is set. The angle features and distance features are selected to make the feature points more prominent. The posture is easy to identify in complex situations. The final posture to be measured is matched with the template posture. The idea of matching calculation is to calculate the real-time posture characteristics, and compare the calculation results with the set type data and thresholds successfully. The significance of the model matching method using one-shot learning small sample learning is that small sample learning does not require a large number of samples and only a small number of samples for training. The advantages of a small number of samples are as follow: 1) The pose features can be accurately located during training; 2) Irrelevant data is discarded; 3) The calculation amount is reduced; 4) The recognition speed and accuracy are improved in the attitude matching process; 5) The anti-interference ability is stronger. The design of this paper is applied to option selection. Gesture recognition replaces the physical device and makes a defined gesture to complete the selection of options. Experimental results show that the method can detect and recognize human poses within the defined threshold in real-time, which improves the accuracy and robustness of recognition.

Li G Y, Li C G, Wang W J, *et al.* Research on multi-feature human pose model recognition based on one-shot learning[J]. *Opto-Electron Eng*, 2021, **48**(2): 200099; DOI: 10.12086/oee.2021.200099

Foundation item: Youth Fund for Science and Technology Research in Colleges and Universities of Hebei Province (2011139) and Natural Science Foundation of Hebei Province (F2012203111)

\* E-mail: 1498578260@qq.com