DOI: 10.19650/j.cnki.cjsi.J2210027

接触式交互感知的人体三维坐姿姿态估计*

周佳裕,蔡晋辉,章 乐,李立新,李晓宇
(中国计量大学计量测试工程学院 杭州 310018)

摘 要:针对视觉姿态估计方法受覆盖遮挡等干扰,提出一种基于座椅面压力图像的人体三维坐姿姿态估计方法,建立坐姿时座椅面体压分布与人体三维姿态之间的跨域联系。设计了一套基于压力-视觉的坐姿训练系统,将阵列式压力传感器嵌入在座椅面中感知坐姿变换,利用时间戳实现压力图像和双目视觉图像的同步匹配。采取双边滤波消除压力图像的尖峰噪声;依靠 OpenPose 姿态估计、三角测量等手段从双目视觉图像中提出 19 个三维关键点;为提高姿态估计精度,提出随机梯度下降最小化损失函数的方法来优化三维关键点坐标,并利用 3D 高斯滤波器进一步生成 3D 关键点置信度图。设计一个基于多层卷积神经网络的压力-视觉跨域深度学习模型,以连续的多帧压力图像作为模型输入,包含三维关键点坐标及其置信度图的 3D 姿态估计结果作为监督对模型进行训练。算法依靠椅面上的阵列传感器接触感知坐姿时的压力分布,就能够准确的估计包含 19 个人体关键点的三维坐姿姿态,在验证集上测试,19 个关键点平均误差 9.7 cm。

关键词: OpenPose;坐姿;阵列式压力传感器;卷积神经网络
中图分类号: TP391 TH701 **文献标识码**: A **国家标准学科分类代码**: 520.60

Human 3D sitting pose estimation based on contact interaction perception

Zhou Jiayu, Cai Jinhui, Zhang Le, Li Lixin, Li Xiaoyu
(*College of Metrology & Measurement Engineering, China Jiliang University, Hangzhou 310018, China*)

Abstract: Aiming at the interference of the visual pose estimation method, such as cover and occlusion, a method of estimating human three-dimensional sitting posture based on the seat surface pressure image is proposed. The cross-domain relationship between seat surface pressure distribution and human three-dimensional posture is established. A posture training system based on pressure and vision is designed. The array pressure sensor is embedded in the seat surface to perceive the posture, and the time stamp is used to realize the synchronization of the visual image matching with the binocular camera. Bilateral filtering is used to eliminate the peak noise of pressure images. Nineteen 3D keypoints are extracted from binocular vision images by OpenPose estimation and triangulation. To improve the accuracy of attitude estimation, a stochastic gradient descent method to minimize the loss function is proposed to optimize the coordinates of 3D keypoints. The 3D confidence graph of keypoints is further generated by 3D Gaussian filter. A multi-layer convolutional neural network pressure-vision cross-domain deep learning model is formulated. Continuous multi-frame pressure images are used as input of the model, and 3D pose estimation results of 3D key point coordinates and their confidence graphs are used as supervision. Based on the pressure distribution of the array sensor on the chair surface, the algorithm can accurately estimate the 3D sitting posture including 19 human key points. The average error of 19 key points is 9.7 cm on the verification set.

Keywords: OpenPose; sitting posture; array pressure sensor; convolutional neural network

0 引 言

研究表明一些疾病的诱因与坐姿不端有较大关联,例如驼背、近视、压疮、颈椎类疾病、腰椎类疾病等身体健康疾病[1]。不良坐姿对正在身体发育的青少年非常不利,对长期处于坐立状态下的上班群体的身体伤害也较大。因此,坐姿姿态估计的研究对人机交互、医疗健康、安全驾驶等具有重要意义。

收稿日期:2022-06-28 Received Date: 2022-06-28
*基金项目:国家自然科学基金(62175223)项目资助

传统的姿态估计是从各种场景的图像中估计人体姿势[2-5],然而基于摄像机的姿态估计仍然具有很多挑战性,包括深度传感器或摄像机等视觉传感器需要校准标定,无法直接安装在应用场景中直接使用;视觉传感器易受环境干扰,如覆盖遮挡[6]、环境亮度变化[7]等;对相机使用时的隐私与安全性的担忧。近年也兴起了基于非视觉人体姿态估计的发展,如 Zhao 等[8]利用 WiFi 频率中穿过墙壁并反射到人体的无线电信号来估计人体的二维姿态。基于压力图像进行姿态估压力图像进行人体姿态估计的研究主要有两种:1)人体与传感器存在大部分接触的床上患者姿态[9-12],Casas 等[13]利用患者与床垫表面接触压力分别通过基于哈希内容检索法和 ConvNet 法成功实现患者三维姿态估计,关键点位置平均误差分别为 12.20 和 8.8 cm;2)利用足部压力分布估计各种日常活动的姿态[14]。不仅如此,人体坐姿相比于床上躺姿,人体与传感器的接触面积小,能获取到的压力信息少;相比于人类日常运动,人体的姿势变化幅度小、频率低,时域上压力特征变化不显著。此外,坐姿下,臀部、脚底等多处受力,特征信息较单一区域受力更为复杂,目前的研究尚未实现基于压力的坐姿姿态估计。

基于上述问题,本文提出一种基于阵列式压力传感器的坐姿姿态估计方法。设计利用坐姿时人体和座椅接触面的压力图像作为输入,以基于相机的姿态估计模型作为监督的深度学习模型,经训练后,网络仅需使用压力图像就能实现坐姿姿态估计,实现接触式感知三维坐姿姿态。贡献一种仅依靠局部压力分布特征——座椅与臀部接触面、无需脚底压力分布的坐姿姿态估计方法,不同于相机姿态估计,传感器嵌入在座椅中,抗干扰能力更强,同时姿态估计模型为下游的分类识别提供了条件。

1　理论分析与设计

1.1　总体设计

本文提出利用座椅椅面的压力图像进行坐姿姿态估计的方法,目前尚无开源的数据集,因此搭建采集系统,收集一些基于压力图像的以视觉数据作为标记姿态的数据来训练深度学习网络。总体流程如图 1 所示,搭建压力-视觉同步采集系统,实时传输人体坐姿变换时的压力数据流和视频数据流;PC 端以时间戳同步匹配的方法接收数据,存储一组包含视觉图像和压力图像的数据对;对压力图像进行去噪处理,对视觉图像采用姿态估计等手段得到以三维关键点坐标及其置信度图表征的人体坐姿姿态标签,制作由一组组以三维坐姿姿态为标签的压力图像数据构成的数据集;数据集用于基于卷积神经网络的坐姿姿态估计模型训练与验证;训练得到的坐姿姿态估计模型仅依靠座椅椅面的压力图像就能预测坐姿。

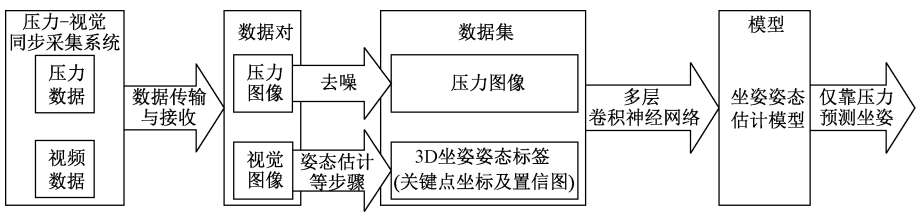

图 1　总体流程

Fig. 1　The overall flow chart

其中,从视觉图像中提取三维坐姿姿态标签的流程如图 2 所示,包括标定和图像识别两部分。标定包括如下两部分:1)采用张正友标定法对相机进行双目标定,确定两个相机之间的相对位姿;2)标定阵列式压力传感器与左相机之间的相对位姿。图像识别如下:首先通过 OpenPose 识别左、右相机的二维图像中的二维人体关键点;其次利用二维关键点坐标和标定的相关参数,采用三角测量法计算三维人体关键点坐标,此坐标是基于左相机的相机坐标系的;然后对于遮挡、未识别的人体关键点利用随机梯度下降最小化损失函数的方法来优化,提高精度,保证数据集的准确性;最后将三维关键点坐标从左相机的相机坐标系转换到基于阵列式压力传感器的物体坐标系,使阵列式压力传感器获取的压力图像与视觉估计的人体三维姿态存在空间对应关系,建立压力-视觉的跨域联系。

1.2　压力-视觉同步采集系统

为采集阵列式压力传感器的压力图像的同时获取人体关键点数据,搭建压力-视觉同步采集系统,采样频率 8 Hz,压力-视觉同步采集系统的系统架构如图 3 所示,包括压力采集端、基于相机的视觉采集端、PC 数据接收端。

将两个海康机器人的工业相机 MV-CA050-12GC 固定在支架上,调整云台角度使人和椅子都进入相机的视场,相机通过 GigE 接口与 PC 相连传输数据,实现视觉采集功能。

如图 4 所示,压力采集端由阵列式压力传感器、数据采集电路组成。在座椅椅面上布置阵列式压力传感器,

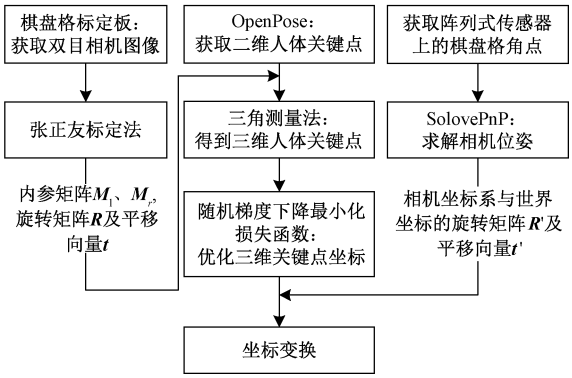

图 2 视觉图像处理流程

Fig. 2 Visual image processing process

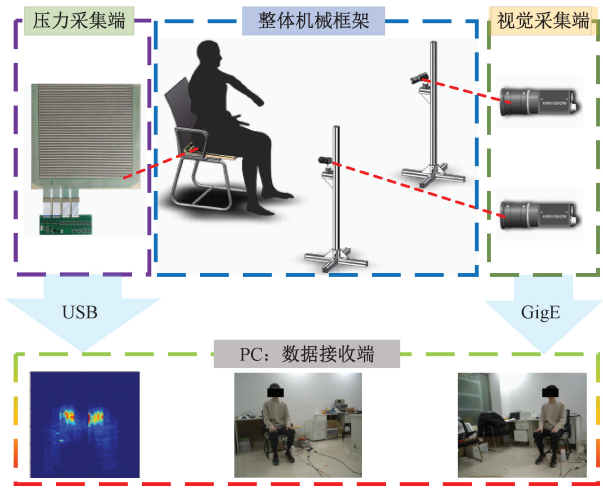

图 3 压力-视觉同步采集系统的系统架构

Fig. 3 System architecture of the pressure-vision synchronous acquisition system

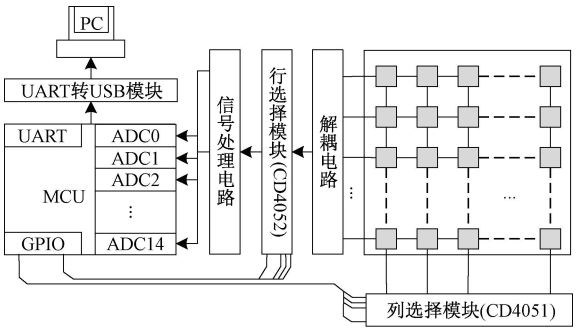

图 4 基于阵列式压力传感器的采集系统

Fig. 4 Acquisition system based on the array pressure sensor

本文采用高分辨率的 MF-6060 阵列式柔性薄膜压力传感器，60×60 的压力传感器阵列形成 3 600 个压力敏感点，压力传感器的阻值随着感应面受到的压力的增加而减小，电阻的倒数与其受到的压力值是一种近似的线性关系。数据采集电路将压力传感器电阻值的变化转换成电压变化，通过行、列选择模块选择阵列中的一个压力单元，通过解耦电路、信号处理电路将电阻值转换成电压值，由单片机进行 ADC 采集，为提高采集速度，采用双 ADC 并行采集。以 8 Hz 的频率发送每组 3 600 个压力点的数据给 PC 端。当坐姿发生变化时，压力成像也会随之改变。

采用多进程并行采集、存储压力图像和视觉图像，保证了采样的实时性。为每一帧数据添加时间戳，校验压力图像、视觉图像的时间戳，匹配时间上误差最小的同步压力-视觉数据对。保证得到两个相机的视觉图像同步，重建得到的 3D 关键点重投影误差小；保证压力图像和视觉图像是同步的，准确反映某一时刻的坐姿。

阵列式压力传感器会记录一些由闪烁噪声、机械和拉伸刺激、温度变化和校准误差引起的噪声[15]。如图 5 所示是不同滤波算法对原始坐姿压力图像处理效果。方框滤波、均值滤波、高斯滤波后压力图像的轮廓边缘也被平滑处理，使得轮廓边缘模糊，丢失了原始图像的轮廓细节，而双边带滤波较好地保留轮廓边缘，但对噪声较大的异常点的处理效果不佳，中值滤波后不仅能很好的保留轮廓边缘，还能有效的滤除尖峰噪声，因此，对压力图像进行中值滤波滤除噪声。

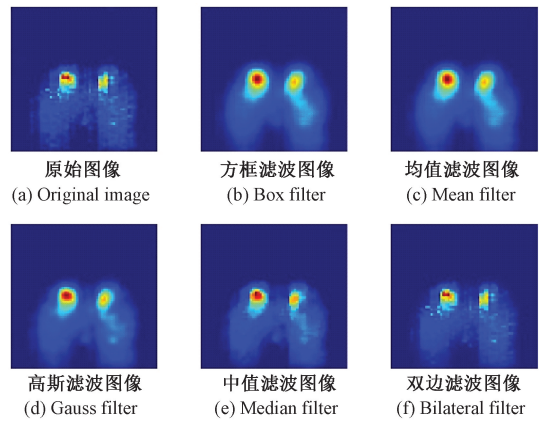

图 5 不同滤波算法对原始坐姿压力图像处理效果

Fig. 5 The processing effect of different filtering algorithms on the original sitting pressure image

1.3 相机的标定与坐标变换

双目标定是根据在同一时间内拍下的双目图片所产生的视差，恢复三维空间物体的三维坐标值的过程，得到成像平面的像点与三维空间该点的对应关系。图 6 所示为采集的 19 组左、右相机标定图片，运用张正友标定法[16]对双目相机进行标定，获取两个相机的内参矩阵 $\boldsymbol{K}_l$ 和 $\boldsymbol{K}_r$、畸变系数以及两相机坐标系之间的旋转矩阵 $\boldsymbol{R}_{3\times3}$ 和平移向量 $\boldsymbol{t}_{3\times1}$，建立二维像素坐标和左相机的相机坐标系的映射关系。

(a) 左相机图像　(b) 右相机图像
(a) Left camera image　(b) Right camera image

图 6　左、右相机的标定图像

Fig. 6　Calibration images of left and right cameras

左相机和右相机的像素坐标系分别到左相机的相机坐标系的转换可表示为:

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \boldsymbol{K}_l \begin{bmatrix} \boldsymbol{E}_{3\times3} & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \boldsymbol{P}^L \begin{bmatrix} \mathrm{X_C} \\ \mathrm{Y_C} \\ \mathrm{Z_C} \\ 1 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \boldsymbol{K}_r \begin{bmatrix} R_{3\times3} & \boldsymbol{t}_{3\times1} \\ \boldsymbol{0} & 1 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \boldsymbol{P}^R \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \tag{2}$$

式中:$\boldsymbol{P}_C[X_C, Y_C, Z_C]$ 为点 P 在左相机的相机坐标系下的坐标;$\boldsymbol{p}_l(u_l, v_l)$,$\boldsymbol{p}_r(u_r, v_r)$ 分别为点 P 在左相机、右相机的像素坐标系中坐标;$\boldsymbol{E}_{3\times3}$ 为单位矩阵;$\boldsymbol{K}_l = \begin{bmatrix} f_{x_l} & 0 & u_{0_l} & 0 \\ 0 & f_{y_l} & v_{0_l} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$,$\boldsymbol{K}_r = \begin{bmatrix} f_{x_r} & 0 & u_{0_r} & 0 \\ 0 & f_{y_r} & v_{0_r} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ 分别为左、右相机的内参矩阵;$\boldsymbol{R}_{3\times3}$ 和 $\boldsymbol{t}_{3\times1}$ 分别为左、右两个相机的相机坐标系之间变换的旋转矩阵和平移向量;$\boldsymbol{P}^L$、$\boldsymbol{P}^R$ 分别是左相机的相机坐标系下的 3D 关键点投影到左、右相机的 2D 图像帧上的投影矩阵。$\boldsymbol{K}_l$、$\boldsymbol{K}_r$、$\boldsymbol{R}_{3\times3}$ 和 $\boldsymbol{t}_{3\times1}$ 这 4 个参数均有双目标定所得。

假设相机-阵列压力传感器系统中,左相机的相机坐标系为 $O_C - X_C Y_C Z_C$,阵列式压力传感器的物体坐标系为 $O_S - X_S Y_S Z_S$,阵列式压力传感器的物体坐标系与左相机的相机坐标系的位姿关系可由旋转矩阵 $\boldsymbol{R}'_{3\times3}$ 和平移向量 $\boldsymbol{t}'_{3\times1}$ 表示。世界点 P 在阵列式压力传感器的物体坐标系中的坐标为 $\boldsymbol{P}_S[X_S, Y_S, Z_S]$,其在左相机的相机坐标系中坐标为 $\boldsymbol{P}_C[X_C, Y_C, Z_C]$,则从左相机的相机坐标系到阵列式压力传感器的物体坐标系的转换关系可表示为:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}'_{3\times3} & \boldsymbol{t}'_{3\times1} \\ \boldsymbol{0} & 1 \end{bmatrix} \begin{bmatrix} X_S \\ Y_S \\ Z_S \\ 1 \end{bmatrix} \tag{3}$$

式中:$\boldsymbol{R}'_{3\times3}$、$\boldsymbol{t}'_{3\times1}$ 分别表示从左相机的相机坐标系到阵列式压力传感器的物体坐标系变换的旋转矩阵和平移向量。

建立阵列式压力传感器的物体坐标系如图 7 所示,以阵列式压力传感器感应区域左上端点为原点,所在平面为 x-y 平面、垂直于该平面的方向为 z 轴。在阵列式压力传感器上粘贴方格边长为 8 cm 的 5×5 棋盘格,识别棋盘格角点 $\boldsymbol{p}_{l_i}$,使用 Solve PnP 迭代法求解 $\boldsymbol{R}'$,$\boldsymbol{t}'$,最小二乘模型为:

$$(\boldsymbol{R}', \boldsymbol{t}') = \operatorname{argmin} \sum_{i=1}^{n} w_i \left\| (\boldsymbol{R}' \boldsymbol{P}_{S_i} + \boldsymbol{t}') - \boldsymbol{P}_{C_i} \right\|^2 \tag{4}$$

式中:$\boldsymbol{P}_{S_i}$ 和 $\boldsymbol{P}_{C_i}$ 分别为一个点在左相机的相机坐标系中的坐标和阵列式压力传感器的物体坐标系中棋盘格角点的坐标,$\boldsymbol{P}_{C_i}$ 根据式(1)计算。

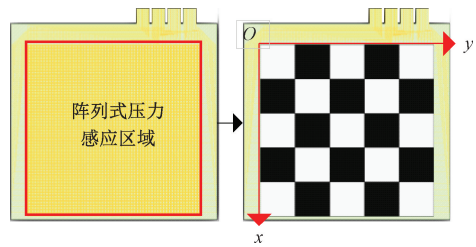

图 7　基于阵列式压力传感器的世界坐标系

Fig. 7　World coordinate system based on the array pressure sensor

1.4　基于双目相机的三维姿态标签的生成

1)基于 OpenPose 的二维姿态估计

从压力图像估计坐姿姿态的一大挑战是缺乏标记数据,采用成熟的视觉姿态估计模型 OpenPose[5] 以解决这一问题,信任得到的姿态估计结果是真实的。OpenPose 是由卡耐基梅隆大学(CMU)感知实验室发布的一个以 OpenCV 和 Caffe 为框架的基于卷积神经网络和监督学习的实时姿态估计方法,该方法采用一种非参数的表达方式,即局部亲和矢量场(part affinity fields, PAF),学习将各个身体部位与图片中的各个人体相关联。其体系结构通过对全局的内容进行编码,从而自下而上地用贪心算法进行解析。其网络结构主要是通过一个连续的预测过程的两个分支来同时学习关键点的定位以及它们之间的关联。OpenPose 的网络结构如图 8 所示,首先通过传统卷积神经网络 VGG19[17] 进行特征提取得到特征图 F,接着将其输入双分支多阶段网络,网络的上支路用于预测关键点置信度图,表征关键点的位置;下支路用于预测关节亲和度,记录关键点之间的位置和方向信息。

OpenPose 识别得到的 25 个关键点,剔除与坐姿变换关联度低的关键点,包括耳朵、大脚趾、脚后跟关键点,降低了后续姿态估计模型的复杂度,选出 19 个关键点,建立人体坐姿姿态模型如图 9 所示。

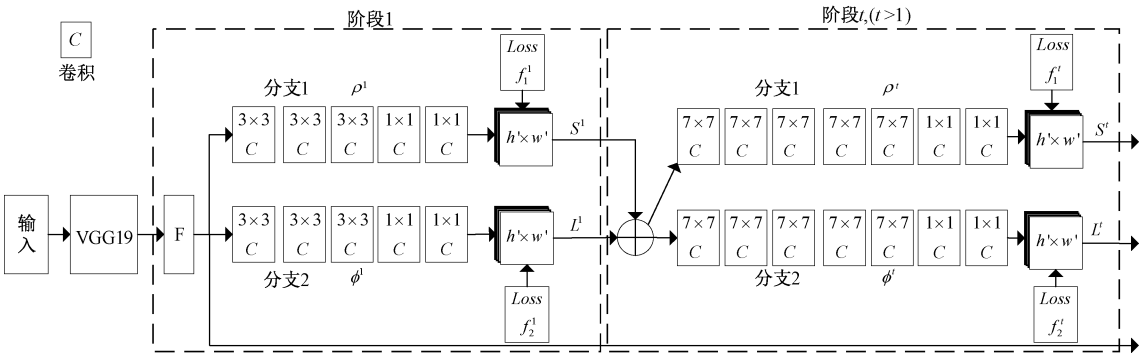

图 8 OpenPose 网络结构

Fig. 8 OpenPose network architecture

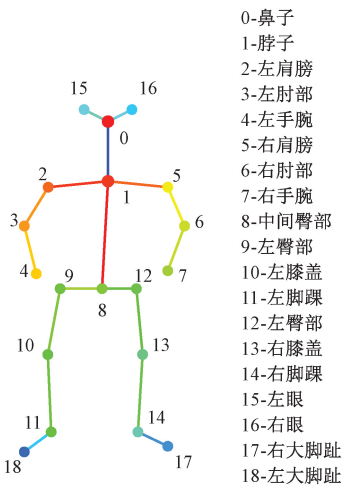

图 9 人体关键点标识

Fig. 9 Human body keypoints identification map

2)人体三维关键点的重建及优化

双目立体定位需要在两个不同视角的图像上找到对应的坐标点[18],图 10(a)所示为对极几何模型,考虑图像 I_l 和 I_r,I_r 到 I_l 的运动为 $\boldsymbol{R}$、$\boldsymbol{t}$。两个相机光心为 O_l 和 O_r,点 P 在 I_l、I_r 中分别对应特征点 $\boldsymbol{p}_l(u_l,v_l)$、$\boldsymbol{p}_r(u_r,v_r)$,依据针孔相机模型的原理,可由式(1)、(2)表示坐标系转换过程。

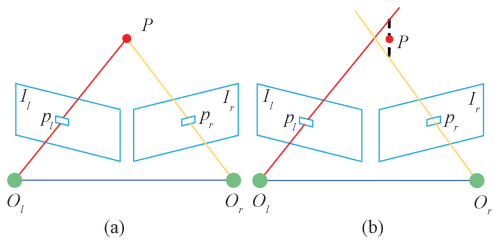

图 10 对极几何模型

Fig. 10 Epipolar geometry

两个视角的图像利用 OpenPose 识别同一个关键点存在一定的偏差,如图 10(b)所示,O_lP 和 O_rP 并不一定相交。利用图像 I_i 和 I_r 匹配的像素点、以及两帧图像的位姿关系 $\boldsymbol{R}_{3\times3}$、$\boldsymbol{t}_{3\times1}$,通过三角测量[19]计算匹配点的 3D 坐标。

从图像中识别出 2D 骨骼关键点并采用三角测量得到 3D 骨骼关键点的过程中,存在两个问题。1)部分动作时的个别人体关键点并不能出现在相机的视场内,无法被 OpenPose 所识别;2)OpenPose 识别骨骼关键点存在一定的误差,对运动模糊时的人体关键点识别准确度不高。

本文提出采用随机梯度下降最小化损失函数的方法来优化骨骼关键点坐标,提高数据集标签的准确性。首先利用三角测量得到三维关键点计算出所有样本的骨骼长度,对于同一个人而言,骨骼长度是恒定的,因此,采用一个人样本集中骨骼长度的中位值作为一个人骨骼真实长度;然后分别计算每个三维人体关键点重投影到左、右相机图像的像素坐标系下的投影误差。损失函数为:

$$L=\sum_{k=1}^{N}\|\boldsymbol{P}^{L}\boldsymbol{P}_{C_k}-\boldsymbol{p}_k^{l}\|+\sum_{k=1}^{N}\|\boldsymbol{P}^{R}\boldsymbol{P}_{C_k}-\boldsymbol{p}_k^{r}\|+\sum_{i=1}^{N-1}\|\hat{K}_i-K_i\| \tag{5}$$

式中:$N=19$,表示人体关键点的个数;$\boldsymbol{P}^L$ 和 $\boldsymbol{P}^R$ 分别表示左相机的相机坐标系下的 3D 关键点投影到左、右相机的 2D 图像帧上的投影矩阵,可由式(1)、(2)计算;$\boldsymbol{P}_{C_k}=\{x_k,y_k,z_k\}$,表示人体关键点 P 位于左相机的相机坐标系下的三维空间坐标;$\boldsymbol{p}_k=(u_k,v_k)$ 表示相应的人体关键点在图像中的像素坐标;$\hat{K}$ 表示一个人骨骼长度的中位值;K 表示三角测量得到的骨骼长度。

人体的坐姿姿态由基于左相机的相机坐标系的 19 个三维人体关键点坐标表征,但相机位置的改变影响坐姿姿态的表征。为了消除相机位置对坐姿姿态表征的影响,同时建立人体三维关键点姿态和压力图像的跨域联系,根据式(4)将三维关键点坐标从左相机的相机坐标系转换到基于阵列式传感器所在位置的物体坐标系,

使得阵列式压力传感器位于基于阵列式压力传感器的物体坐标系的 $x-y$ 平面,即压力图像位于该坐标系的 $x-y$ 平面,同时人体坐姿姿态也是由基于该坐标系下的三维人体关键点坐标表征的。

3D 空间上对关键点位置应用 3D 高斯滤波器进一步生成 3D 关键点置信度映射得到三维关键点的置信度图,如图 11 所示。

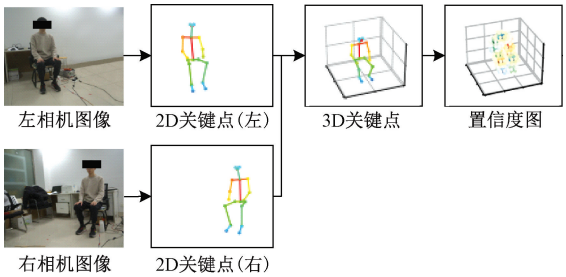

图 11　置信度图生成过程

Fig. 11　Confidence graph generation process

2　实　　验

2.1　实验环境

实验在由 Windows 10 操作系统、Python 3.6.13, PyTorch 1.6.0+cuda 10.1 以及一个 NVIDIA TITAN Xp GPU 组成的服务器上完成。

利用上述的压力-视觉同步采集系统及数据处理方法,实现对 5 个被测人员进行实时动态坐姿采集,每个人在座椅上不停地以不同速率、不同幅度自由地变换坐姿,例如从左倾到右倾、从右倾到正坐等,并对采集的数据进行处理,共收集了 90 000 组同步的压力和视觉数据对样本,平均同步误差 62.5 ms,包括前倾、后倾、左倾、右倾、正坐、驼背、左右转身等坐姿类型。每个样本由 60×60 个压力点的压力图像、19 个基于传感器的物体坐标系的人体关键点三维坐标及其置信度图组成,19 个人体关键点包括头、脖子、肩膀、肘部、腰部、臀部、膝盖、脚踝和大脚趾。

5 个被测人员生理信息如表 1 所示,5 个人的身高、体重分布具有良好的分散性,有助于训练所得模型的泛化。训练集中包含了被测人员 1、2、3、4 的坐姿数据,共 70 000 个样本;验证集 1 中包含了被测人员 1、2、3、4 的坐姿数据,共 15 000 个样本;验证集 2 中包含了被测人员 5 的坐姿数据,共 5 000 个样本。

2.2　基于卷积神经网络的姿态估计模型

在 Luo 的模型[14]基础上修改,模型如图 12 所示,以时间上连续采样的 M 帧压力图像作为模型的输入,从视觉图像中获取的 19 个人体关键点的三维坐标及其三维置信度图作为监督来训练模型,输出与中间帧相对应的 19 个人体三维关键点的置信度图。压力图像位于 $x-y$ 平面上,与人体关键点的坐标有很好的空间对应关系,将时间上连续的 M 帧二维的压力图像,经过多层 2D 卷积处理后,为了回归得到 3D 空间中人体关键点坐标,在网络中添加一个新的维度,从二维空间扩展至三维空间,同时添加一个通道来表征 z 轴的高度,然后利用多层 3D 卷积对特征进行处理,预测 19 个关键点的三维关键点置信度图,该模型的每层网络结构如表 2 所示。

表 1　被测人员身高和体重信息

Table 1　Height and weight information of the subjects

被测人员	身高/cm	体重/kg
1	184	78
2	180	90
3	179	55
4	158	51
5	166	55

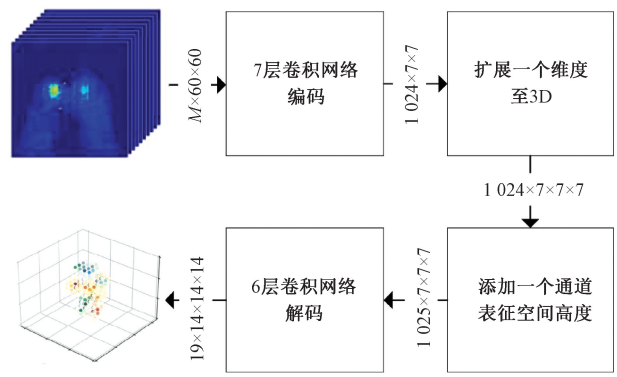

图 12　三维人体姿态估计模型

Fig. 12　3D human pose estimation model

采用 Adam 优化器最小化预测关键点置信度图和真实情况下的关键点置信度图之间的均方误差(MSE)来优化模型,同时,添加了两部分损失项。1)为了保证输出的骨骼长度符合正常人体,计算骨骼长度使其位于样本对象人体骨骼长度的 3%~97%;2)为准确关联坐姿变化,加入能充分体现坐姿变换的肢体间角度误差。最后使用 Softmax 将置信度图转为关键点三维坐标。损失函数的定义为:

$$Loss=\frac{1}{N}\sum_{i=1}^{N}|H_i-\hat{H}_i|+\frac{1}{N-1}\sum_{j=1}^{N-1}L_j+\frac{1}{N2}\sum_{i=1}^{N2}|\theta_i-\hat{\theta}_i| \tag{6}$$

式中:H_i 和 $\hat{H}_i$ 分别表示真实情况下的和模型预测得到的关键点置信度图;θ_i 和 $\hat{\theta}_i$ 分别表示真实情况下的和模

表 2 三维人体姿态估计模型各层结构

Table 2 The structure of each layer of the 3D human pose estimation model

层数	结构	输出维度
1	Conv2D, LeakyReLU, BatchNorm	32×60×60
2	Conv2D, LeakyReLU, BatchNorm, MaxPool2D	64×60×60
3	Conv2D, LeakyReLU, BatchNorm	128×30×30
4	Conv2D, LeakyReLU, BatchNorm, MaxPool2D	256×15×15
5	Conv2D, LeakyReLU, BatchNorm	512×15×15
6	Conv2D, LeakyReLU, BatchNorm	1 024×14×14
7	Conv2D, LeakyReLU, BatchNorm, MaxPool2D	1 024×7×7
8	扩展一个维度	1 024×7×7×7
9	添加一个通道	1 025×7×7×7
10	Conv3D, LeakyReLU, BatchNorm3D	1 025×7×7×7
11	Conv3D, LeakyReLU, BatchNorm3D	512×7×7×7
12	ConvTranspose3D, LeakyReLU, BatchNorm3D	256×7×7×7
13	Conv3D, LeakyReLU, BatchNorm3D	128×14×14×14
14	Conv3D, LeakyReLU, BatchNorm3D	64×14×14×14
15	Conv3D, Sigmoid	19×14×14×14

型预测得到的相连接骨骼所成角度 L_j 表示第 j 个骨骼长度损失，定义为：

$$L_j = \begin{cases} K_j^{\min} - \hat{K}_j, & \hat{K}_j < K_j^{\min} \\ \hat{K}_j - K_j^{\max}, & \hat{K}_j > K_j^{\max} \\ 0, & \text{其他} \end{cases} \tag{7}$$

式中：$\hat{K}_j$ 表示模型预测得到的骨骼长度；$K_j^{\min}$ 和 $K_j^{\max}$ 分别表示训练集中样本的 3% 和 97% 的骨骼长度。

2.3 数据分析

用 70 000 组数据对的训练集对基于卷积神经网络的姿态估计模型进行训练，并使用独立 15 000 组数据对的验证集 1 对训练得到的模型进行验证。

采用欧氏距离计算关键点位置误差来评价模型预测得到的三维人体坐姿与座椅上真实的人体坐姿姿态之间的误差，如图 13 所示，是采用时间上连续的 10 帧压力分布图作为输入训练得到模型在测试集上的验证结果，对于与座椅上阵列式压力传感器距离越接近的关键点，位置误差越小，如臀部、膝盖等；对于距离躯干主轴越近的关键点，位置误差也越小，如脖子、肩膀和膝盖等；对于距离躯干主轴越远的关键点，位置误差越大，如手腕、肘部等。19 个关键点的总体平均误差为 9.7 cm。

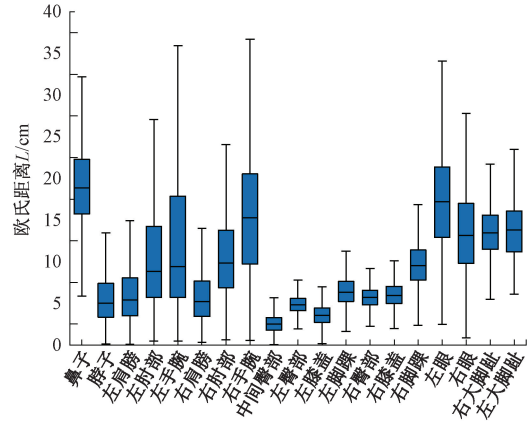

图 13 不同关键点的估计误差

Fig. 13 Estimation error at different keypoints

图 14 所示是上述训练所得的姿态估计模型对一些坐姿姿态的预测结果，可以看出，该模型对各种坐姿都有良好的预测能力，对靠近身体主干和靠近传感器的关键点的预测能力较强，如脖子、肩膀、臀部等关键点；相反对较远的关键点的预测精度稍差，如头部、手腕等；总体而言，完全可以预测出坐姿的整体趋势。

为测试模型的鲁棒性，使用模型训练时尚未出现的被测人员 5 的坐姿数据来测试模型性能，即验证集 2，结果如图 15 所示，关键点位置误差略大于前者，平均误差 15.3 cm。可以看出，模型具有一定的泛化能力，在全新的个体上也能实现坐姿姿态估计。

图 16 所示是不同输入帧数下的模型性能，从单帧输入开始，随着输入帧数的增加，模型的损失值逐渐减小，至输入帧为 10 帧时，损失值最小，模型性能表现最佳，继续增加输入帧数，模型的损失值迅速增加，模型的性能下降。

3 实际应用

为了展示压力图像预测得到的坐姿骨架特征可以用于各种识别任务，因此，利用本文基于压力图像的坐姿姿态估计输出的骨架进行分类识别算法验证。采集坐姿时压力图像和视觉图像的数据对，共 5 000 组；压力图像经本文的姿态估计模型预测得到 19 个人体三维关键点，根据视觉图像人工标注坐姿类型，确定坐姿类型包括前倾、右倾、后倾、左倾、正坐 5 种坐姿；采用随机森林算法进行坐姿分类，70% 用于训练，30% 用测试。

使用随机森林分类算法对本文姿态估计模型输出的特征（坐姿下的三维关键点坐标）进行坐姿分类，测试集

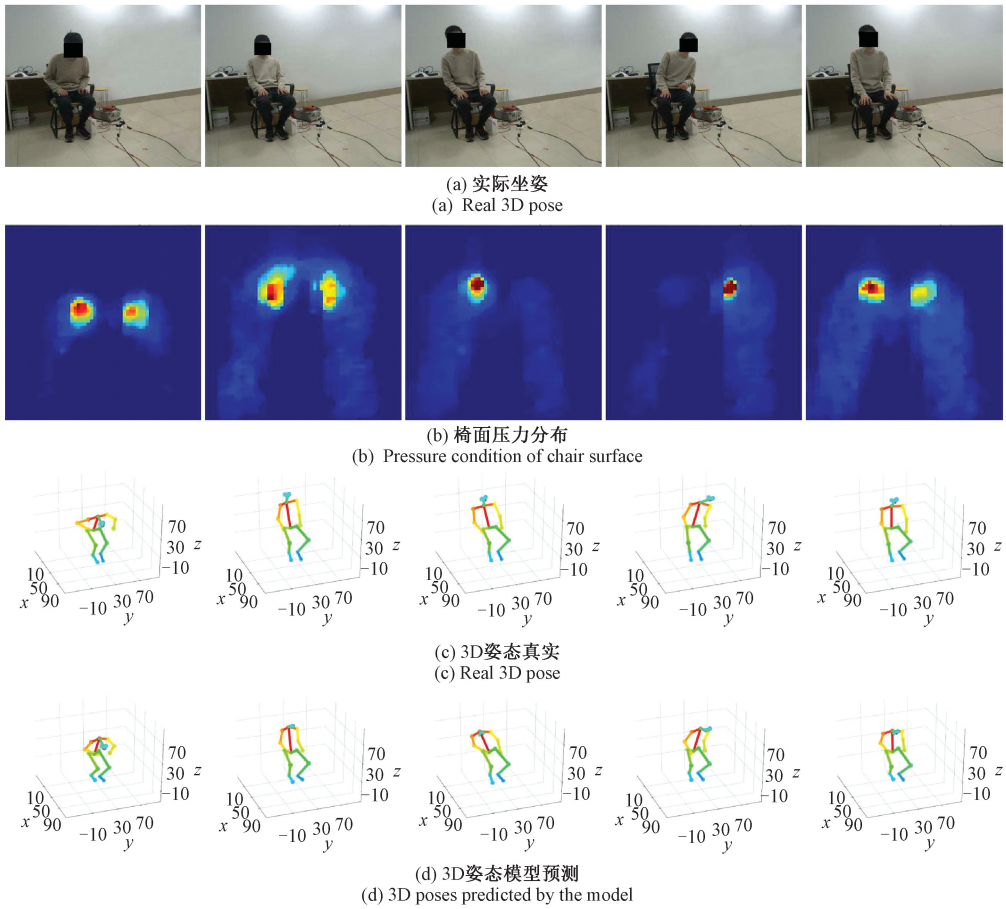

图 14　时间步长 10 帧的三维坐姿姿态估计结果

Fig. 14　Estimation results of 3D sitting posture with a time step of 10 frames

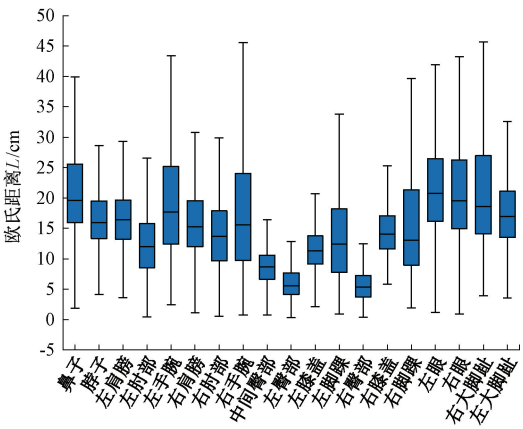

图 15　模型在验证集 2 上的姿态估计误差

Fig. 15　The attitude estimation error of the model on validation set 2

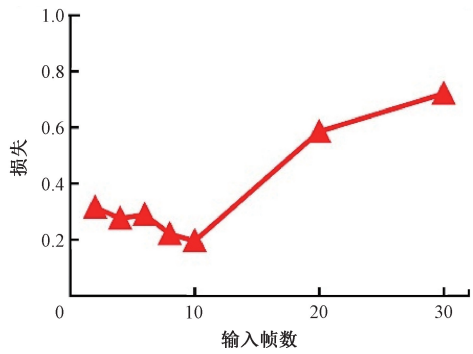

图 16　输入帧数 frame 与损失函数 loss 关系

Fig. 16　The input frame number and the relationship between the loss function loss

准确率达 91.1%,如图 17 所示。不仅说明了本文依靠压力图像预测坐姿时的三维关键点姿态的准确性,还说明了根据人体三维关键点坐标进行坐姿分类具备可行性。再次证明,本文依靠坐姿时椅面压力图像预测人体坐姿姿态(3D 关键点)的方法具有一定的实用价值,本文姿态估计模型对下游的分类算法研究具有一定的帮助。

4　结　　论

安装不便、易受环境干扰和隐私安全等问题是人体姿态估计和许多视觉任务中都存在的问题。本文实现了基于阵列式压力传感器的三维坐姿姿态估计。设计了一

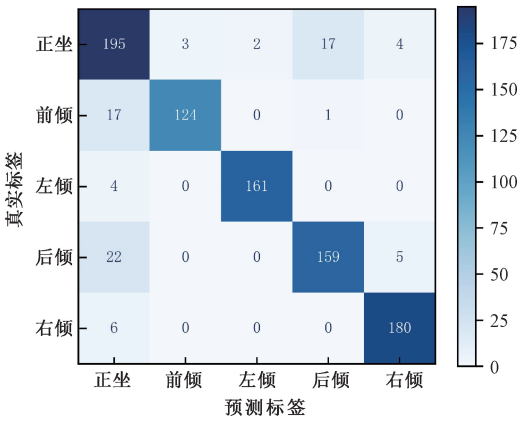

图 17 随机森林算法对坐姿分类的混淆矩阵

Fig. 17 Confusion matrix for sitting posture classification by random forest algorithm

套基于压力-视觉的坐姿训练系统,实现压力图像和视觉图像的同步采集;利用成熟的 OpenPose 姿态估计模型等手段从视觉图像中获取基于阵列式压力传感器坐标系下的人体三维关键点,建立压力图像与三维关键点的跨域联系;进一步以多帧压力图像作输入,视觉的姿态估计结果作为监督训练模型,本模型仅依靠连续的多帧压力图像便能推断出包含 19 个关键点的三维坐姿姿态。1)本文设计的姿态估计模型仅需与人体无缝接触的椅面压力图像就能估计出人体的 3D 坐姿姿态,解决了视觉姿态估计方法总是受到办公桌等环境的遮挡,无法获取全身的人体姿态;2)利用人体三维关键点的坐标计算关节角度实现自动化标注压力图像的坐姿类型标签,解决了阵列式薄膜压力传感器应用于深度学习的坐姿检测研究时人工标注坐姿类型速度慢、效率低这一痛点;3)建立了压力图像与视觉图像的跨域联系,仅依靠坐姿时人与外界局部接触区域的力学特征——臀部与椅面,实现从 2D 信息中恢复 3D 信息,实现压力图像到人体姿态的估计,区别于躺姿仅与床面接触,日常跑步等仅脚底与地面接触;4)该坐姿姿态估计方法不同于视觉方法需要额外安装摄像头,只需在椅面嵌入阵列式压力传感器,使用更加方便。

参考文献

[1] 郑泽铭. 人的坐姿检测方法及行为劝导研究[D]. 杭州:浙江大学,2013.
ZHENG Z M. Sitting posture sensing and healthy behavior persuation[D]. Hangzhou: Zhejiang University, 2013.

[2] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C]. CoRR, 2016.

[3] 孟琭,高恒上,张含光,等. 基于全连接神经网络的三维人体姿态估计[J]. 仪器仪表学报,2020,41(10):165-177.
MENG L, GAO H SH, ZHANG H G, et al. Three-dimensional human pose estimation based on the fully connected neural network [J]. Chinese Journal of Scientific Instrument, 2020,41(10):165-177.

[4] PAPANDREOU G, ZHU T, CHEN L, et al. PersonLab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model[C]. Computer Vision-ECCV, 2018.

[5] CAO Z, SIMON T, WEI S, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]. 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), 2017:1302-1310.

[6] ACHILLES F, ICHIM A E, COSKUN H, et al. Patient MoCap: Human pose estimation under blanket occlusion for hospital monitoring applications [C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2016: 491-499.

[7] LIU S, YIN Y, OSTADABBAS S. In-bed pose estimation: Deep learning with shallow dataset[J]. IEEE Journal of Translational Engineering in Health and Medicine, 2019,7:4900112.

[8] ZHAO M, LI T, ABU ALSHEIKH M, et al. Through-wall human pose estimation using radio signals[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018:7356-7365.

[9] CLEVER H M, ERICKSON Z, KAPUSTA A, et al. Bodies at rest: 3D human pose and shape estimation from a pressure image using synthetic data[J]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020:6214-6223.

[10] CLEVER H M, KAPUSTA A, PARK D, et al. 3D human pose estimation on a configurable bed from a pressure image [J]. Computer Science, 2018, DOI: 10.48550/arXiv.1804.07873.

[11] BELAGIANNIS V, ZISSERMAN A. Recurrent human pose estimation[C]. CoRR, 2016.

[12] GRIMM R, BAUER S, SUKKAU J, et al. Markerless estimation of patient orientation, posture and pose using

range and pressure imaging[J]. International Journal of Computer Assisted Radiology & Surgery, 2012, 7(6): 921-929.

[13] CASAS L, NAVAB N, DEMIRCI S. Patient 3D body pose estimation from pressure imaging[J]. International Journal of Computer Assisted Radiology and Surgery, 2018,14(3):517-524.

[14] LUO Y, LI Y, FOSHEY M, et al. Intelligent carpet: Inferring 3D human pose from tactile signals[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:11250-11260.

[15] ISHAC K, SUZUKI K. LifeChair: A conductive fabric sensor-based smart cushion for actively shaping sitting posture[J]. Sensors, 2018,18(7): 2261.

[16] ZHANG Z. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11):1330-1334.

[17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]. CoRR, 2014.

[18] 常晓龙. 基于深度学习的人体运动目标实时检测定位与动作识别[D]. 杭州:浙江理工大学,2019.
CHANG X L. Real-time detection,positioning and motion recognition of human moving objects based on deep learning[D]. Hangzhou: Zhejiang Sci-Tech University, 2019.

[19] HARTLEY R, ZISSERMAN A. Multiple view geometry in computer vision[M]. 2nd ed. Cambridge, UK: Cambridge University Press, 2004.

作者简介

周佳裕,2020 年于中国计量大学获得学士学位,现为中国计量大学硕士研究生,主要研究方向为坐姿监测。

E-mail: zhoujiayu_cjlu@163.com

Zhou Jiayu received his B. Sc. degree from China Jiliang University in 2020. He is currently a M. Sc. candidate at China Jiliang University. His main research interests include sitting position monitoring.

蔡晋辉(通信作者),2005 年于浙江大学获得博士学位,2005~2007 年在浙江大学自动化仪表所从事博士后研究工作,现为中国计量大学教授,主要研究方向为检测技术与信号处理。

E-mail: caijinhui@cjlu.edu.cn

Cai Jinhui (Corresponding author) received his Ph. D. degree from Zhejiang University in 2005. He worked as a postdoctoral research fellow in Institute of Process Automation Instrumentation at Zhejiang University from 2005 to 2007. He is currently a professor at China Jiliang University. His main research interests include detection technology and signal processing.