

基于人体骨架信息的行为检测系统设计

雷振轩

(西安建筑科技大学 信息与控制工程学院, 西安 710055)

摘要: 为了解决现有行为检测系统中依赖惯性传感器、检测结果不够准确的问题, 设计了基于人体骨架信息的行为检测系统; 系统采用 Jetson Nano 人工智能计算设备作为主控模块, 结合图像采集模块、显示模块和以 Atmega328 单片机为主的报警模块构成; 系统利用图像采集模块采集行为视频信息, 通过主控模块中的行为检测器对视频中人体行为进行检测, 报警模块通过串口接收检测结果并对危险行为进行预警; 同时, 利用人体骨架的关节空间运动幅度、肢体关联差异, 建立了关节帧间位移矢量和骨骼夹角变化的关节行为模型, 再借助长短时记忆网络提取行为特征, 并训练实时行为检测器; 经实验测试, 该系统能够有效检测常见的人体行为并对危险行为类别进行报警提示。

关键词: 行为检测; 关节行为模型; 长短时记忆网络; Jetson Nano; 单片机

Design of Action Detection System Based on Human Skeleton Information

LEI Zhenxuan

(College of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: In order to solve the problems that the existing action detection system relies on inertial sensors and the detection results are not accurate enough, an action detection system based on human skeleton information is designed. The system uses Jetson Nano artificial intelligence computing equipment as the main control module, the auxiliary modules consist of the image acquisition module, display module and alarm module, alarm module based on Atmega328P microcontroller. The system uses the image acquisition module to collect the action video information, detects human action in the video through the action detector in the main control module, and the alarm module receives the detection results through the serial port and gives early warning of dangerous actions. At the same time, using the joint space motion range of the human skeleton and the difference of limb associations, a joint action model for the displacement vector between joint frames and changes in the bone angle is established, and then the long short term memory (LSTM) network is used to extract the action features and train the real time action detector. After the experimental test, the system can effectively detect common human action and give alarms to dangerous action categories.

Keywords: action detection; joints action model; LSTM network; jetson nano; micro control unit

0 引言

行为检测作为计算机领域中的研究热点之一, 在智能监控^[1]、人机交互^[2]等领域有着广泛的应用前景。行为检测可用于室外公共场合人员的危险、暴力行为检测监控与报警, 室内场景下可用于老人和儿童的安全监控^[3-4]。

现有的行为检测技术中有部分是通过红外传感器来感知人体在空间中的位置信息, 并结合场景中的物品位置信息综合判断人体当前的行为状态^[5-6]。这种方法可以大致判断人体处于何种运动状态, 但是由于其传感器工作原理的限制, 只能感知人体空间移动状态, 而对于细致的肢体动作并不能做出精确的检测。随着移动设备和可穿戴设备的发展, 李鹏欢^[7]和黄龙等人^[8]采用加速度传感器获取人体运动信息, 并结合相关的分析算法对获取的运动信息进行分析得出人体行为; 石磊鑫等人^[9]设计了基于可穿戴设备的行为识别系统, 通过将可穿戴设备上的加速度传感器、心

率传感器检测到的人体信息通过蓝牙发送到手机等处理设备综合判断人体行为。相比基于红外传感器的检测方法, 惯性传感器的应用能够更加精细的获取人体的运动信息, 但是通过惯性信号也只能检测一些行走、跑步等简单的行为。这种基于惯性传感器的行为检测方法并不适用于公共场合的群体行为监控场景。

近年来, 卷积神经网络 (CNN, convolutional neural networks)^[10]在图像处理、目标检测方面显示出了极大的优势。Simonyan^[11]、Zhu^[12]和 Zhao 等人^[13]采用双流卷积神经网络 (Two-stream CNN, two-stream convolutional neural network) 来解决行为检测问题, 即提取单帧 RGB 图像中人体行为的空间外观特征, 形成空间流, 再将包含人体行为的连续光流图输入到深度卷积神经网络中, 提取时间轨迹特征, 形成时间流, 最后将时间特征和空间特征融合完成行为检测。这种空间流与时间流的特征结合模式取得了较好的识别准确率。但直接卷积提取空间特征也会引入背

收稿日期: 2022-06-13; 修回日期: 2022-07-18。

基金项目: 陕西省重点研发计划社会发展领域项(2022SF-242)。

作者简介: 雷振轩(1995-), 男, 陕西宝鸡人, 硕士研究生, 主要从事图像处理方向的研究。

引用格式: 雷振轩. 基于人体骨架信息的行为检测系统设计[J]. 计算机测量与控制, 2023, 31(1): 30-35, 44.

景等冗余信息，加之光流特征对光线变化敏感，同样存在复杂背景下的不稳定性。为了更好地利用视频中的时序信息，一些学者^[14-16]采用三维卷积网络（C3D, convolutional 3D）在时间和空间维度上对连续视频帧进行卷积操作，以捕捉视频流中的人体运动信息，学习行为特征。也有人利用长短时记忆网络（LSTM, long short term memory）在时间序列处理上的优势，将它与卷积神经网络相结合构造时空特征融合的行为检测方法^[17-18]，取得了较好的效果。

分析现有行为检测技术的长处和问题，也观察到骨骼和关节在人体运动过程中的行为表达直观性，以及对背景、光线等外部因素干扰的鲁棒性，设计了基于人体骨架信息的行为检测系统。系统以 NVIDIA 公司推出的 Jetson Nano 作为硬件平台的核心，搭建具备行为画面采集、行为类别检测、危险行为报警和行为画面显示功能的硬件系统平台；另外在人体关节坐标基础上构建关节位移矢量和骨骼夹角变化的关节行为模型，借助长短时记忆网络 LSTM 提取行为特征，实现行为检测功能。

1 系统结构及原理

系统主要由图像采集模块、主控模块、显示模块和报警模块组成，系统结构如图 1。其中图像采集负责采集当前行为画面并输入主控模块；主控模块对行为画面进行人体关节提取以及行为检测，然后将检测结果通过串口发送至显示模块和报警模块；显示模块将人体骨架姿态和检测结果与原画面叠加实时显示；报警模块由 Atmega328P 单片机、蜂鸣器和发光二极管组成，对接收到的行为类别进行判断，如果当前行为为危险行为或者有潜在危险，则报警模块报警。

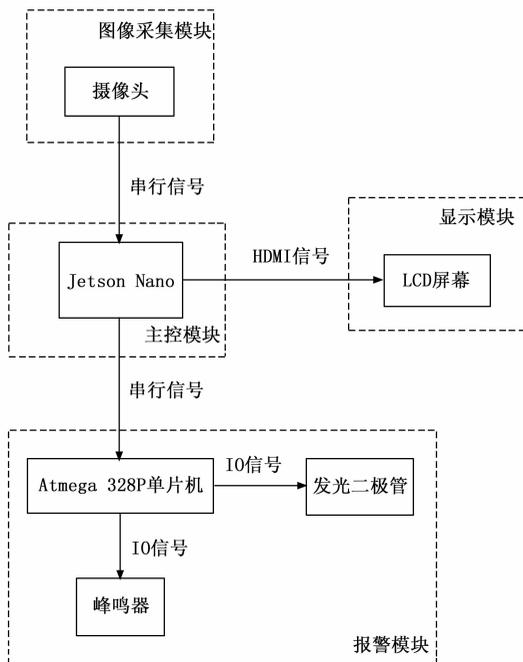


图 1 行为检测系统结构图

2 系统硬件设计

2.1 图像采集模块

作为行为检测系统工作流程的第一步，需要图像采集模块对人体运动画面进行拍摄采集。传统的摄像头大多由数据总线、时钟总线和同步信号线组成，这一类摄像头的接口需要的数据线数量较多，内部的逻辑结构也较为复杂；同时对于水平同步信号、垂直同步信号和时钟信号的同步性要就较高，这使得摄像头数据的输出端和画面信号的接收端均需要具备较好的性能以满足信号的传输。除此之外，在摄像头数据的传输过程中，所传输到信号比较容易受到外界的干扰，这就难以保证在持续性传输画面过程中的画面质量以及传输画面的速率。而 CSI 摄像头在对画面像素信息传输的过程中采用数据差分信号，保证了较好的稳定性，而且该摄像头能够仅使用一组数据信号线和差分时钟线完成数据传输，降低了接口线的使用数量。

考虑到嵌入式系统在图像处理性能上的局限性，在对比了 CSI 摄像头和传统摄像头后，选择占用计算资源较少的 CSI 摄像头来完成图像采集工作，在满足相应的传输速率的同时，也具有更高的传输质量。CSI 摄像头参数如表 1 所示。该摄像头除了占用计算资源少的优点外，还具备体积小、分辨率高、感光器件尺寸大、画面畸变小的优点。

表 1 摄像头参数

项目	参数
感光芯片	索尼 IMX219
分辨率	800 W(3 280 * 2 464)
CMOS 尺寸	1/4 英寸
光圈	2.0
焦距	2.96 mm
畸变	<1%

2.2 主控模块

主控模块主要有以下 3 个功能：（1）接收图像采集模块发送的视频数据。（2）运行人体行为检测器对接收的画面提取关节信息并检测行为类别。（3）通过串口将行为检测结果发送至显示模块和报警模块。

主控模块作为整个系统的核心单元，在运行人体行为检测器对接收的画面提取关节信息并检测行为类别时，需要具备较强的数据运算处理能力。Jetson Nano 是 NVIDIA 公司推出的一款性能强大、体积小的人工智能计算设备，采用最高频率为 1.43 GHz 的 64 位四核 ARM 处理器、128 核 NVIDIA Maxwell GPU、4G 运行内存，支持 5 V DC 输入^[19]。Jetson Nano 能够运行多种算法和人工智能框架，为各种复杂的深度神经网络模型提供实时计算和推理，可以实现图像分类、目标检测、人脸检测与识别、语音处理等功能，其较强的运算能力，比较适合处理摄像头拍摄的图像数据以及运行人体行为检测网络模型。因此，系统中采用 Jetson Nano 作为主控模块，除了运行训练好的人体行为检测模型以外，还兼顾对其余模块发送控制指令和检测结

果信息的功能。

2.3 报警模块

报警模块负责接收主控模块发送的行为类别信息，并判断行为是否属于危险行为，同时发送控制信号控制蜂鸣器、发光二极管进行声光报警。

报警模块采用 Atmega328P 作为控制器，通过串口与主控模块进行通信，并输出高低电平信号给蜂鸣器和发光二极管从而实现报警功能。Atmega328P 是一个高性能、低功耗的 8 位微控制器以 AVR 作为核心处理器，闪存容量为 32 KB，时钟频率为 20 MHz，支持 5 V 输入，具备 I²C、SPI、USART 等接口。以 Atmega328P 单片机为核心设计报警模块及其外围电路，电路原理图如图 2。将单片机 XTAL1 和 XTAL2 外接一个 16 MHz 晶振和两个 20 pF 电容组成外部振荡电路，为单片机内部提供稳定的时钟信号流；同时，将两个发光二极管以共阳极的接法分别连接至单片机的 IO 接口 PD2 和 PD3，通过 IO 接口高低电平来控制发光二极管亮灭，IO 接口置低电平，发光二极管亮，IO 接口置高电平，发光二极管灭。

另外，对比了有源蜂鸣器和无源蜂鸣器的特点，有源蜂鸣器内部含有震荡源，通电后只能发出一种类型的声音，无源蜂鸣器内部不含震荡源，需要通过改变输入信号即可自由控制声音频率；由于无源蜂鸣器的声音频率可控，因此采用无源蜂鸣器作为报警器件，将蜂鸣器控制端连接至 Atmega328P 单片机 IO 接口的 PD4 端，通过单片机输出 1.5~2.5 kHz 频率的高低电平信号即可控制声音频率变化。

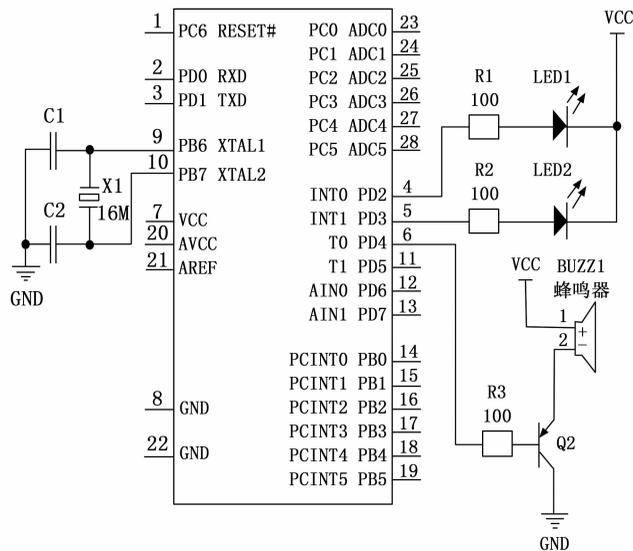


图 2 报警模块电路原理图

3 行为检测方法设计

行为检测方法作为系统软件的核心，主要包括人体关节节点提取、关节行为模型构建和行为特征提取与分类三大部分。

3.1 人体关节节点提取

通过 Jetson Nano 控制 CSI 摄像头并实时读取摄像头拍摄到的视频画面，在此视频画面基础上首先需要提取目标人体的关节信息。随着人体姿态估计算法的不断成熟，现在已有从原始视频画面中提取人体关节信息的深度神经网络模型，例如 OpenPose、DensePose、Mask R-CNN 等，但考虑到 Jetson Nano 模块的计算资源有限，同时为了便于后续在此基础上扩展其他目标检测功能，因此选用兼具目标检测、实例分割和人体姿态估计功能的 Mask R-CNN 从视频中获取关节信息，后续再构建关节行为模型，支持人体行为识别。

在 Mask R-CNN 网络中，掩码分支网络 mask branch 是用于预测并生成目标分割掩码的分支网络，通过双线性插值的方法来保证输入与输出的像素一一对应。在进行特征提取后，mask branch 对每个感兴趣区域进行分类预测，并根据其分类结果对预测的目标使用不同颜色的掩膜进行渲染可视化，作为对目标整体轮廓的描述，来实现像素级的实例分割。这种像素级的实例分割方式也可用于检测并分割人体的关节点，将每一个关节点的位置设为一个二进制单点，每一个 mask 中若有对应的关节点标为 1，反之则为 0，即热图中仅保留概率值为 1 的关节点像素。然后利用全卷积网络作为姿态检测器对感兴趣区域内的人体特征图进行卷积池化和反卷积，得到高分辨的特征输出，同时利用 ROI Align 进行像素到像素的对齐，级联边界框回归与关节点分类。每一个感兴趣区域的 ROI Align 都对应 $K \times m^2$ 维度的输出，其中 K 代表 mask 的类别个数， m^2 代表池化分辨率，在这里仅指 $m \times m$ 的二进制掩码。对于一个人体实例的 18 个关节点，输出 18 个掩码特征图，18 个掩码相互独立，并对每个像素的误差值进行交叉熵损失迭代，最终使关节点像素的输出为 1，其余输出为 0，从而实现对于人体关节点的检测提取。

Mask R-CNN 网络输出的关节点为 COCO 格式，包括 18 个点的坐标位置数据，然而这 18 个点中的左眼、右眼、左耳、右耳这 4 个点在人体运动过程中相对位置不会发生变化同时与人体行为的相关性较小，而且这 4 个点的空间位置相近，在后续的行为检测中容易对识别产生干扰。因此从中剔除这 4 个点，仅保留能够代表肢体运动的 14 个关节点数据，作为后续关节行为模型的输入信息。

3.2 关节行为模型构建

观察人体行为中的各关节运动规律，可发现在一个行为中并非所有关节同时运动，每个关节的运动幅度大小不一，这就意味着各关节的运动信息与当前人体行为的相关性存在差异。因此，将各关节的位置坐标序列送入 LSTM 网络，即可学习行为中前后帧之间的关节时序特征；但是这种直接采用原始关节点位置坐标的方式不能明显的表征人体运动过程总各个关节点的运动幅度差异，为了进一步描述各关节点运动幅度的差异性，针对人体的每个关节点构造了当前帧与上一帧之间的关节点位移矢量模型，通过

同一关节在前后两帧中的位置坐标，计算出该点在前后两帧之间的位移矢量，以每个关节的帧间位移矢量大小来表征该关节的与运动幅度。

除此之外，一些关节在人体行为过程中并没有发生明显的空间位置移动，但是仍对于当前的行为具有较大的影响。例如人体在执行挥手动作时，除了肘部、手部两个关节与当前行为动作的相关性较高以外，该侧肩部的运动也对当前行为动作具有较大的影响，但是肩部关节在挥手动作中并没有产生较大的空间位移而是发生了旋转。这种情况下，帧间位移矢量模型并不能描述旋转点与当前行为的相关性大小，然而关节发生旋转时该点与其两侧相邻关节构成的夹角也会发生明显变化，因此，考虑以除过肢体末端以外的其余关节与其两侧相邻的两个关节构成的夹角来表征该点的旋转情况，具体构成夹角的关节组合包括：左右手臂的手、肘、肩；左右腿部的足、膝、胯；左侧肘、肩和颈；右侧肘、肩和颈；左侧膝、胯和颈；右侧膝、胯和颈；头、颈、左肩；头、颈、右肩；头、颈、左胯；头、颈、右胯；左肩、颈、左胯；右肩、颈、右胯；共计 14 个夹角数据。然后，计算每个夹角在前后帧间的夹角变化量，得到一个帧间骨骼夹角变化模型。

最后将原始关节坐标信息、帧间位移矢量模型的输出和帧间骨骼夹角变化模型的输出这三路信息融合提升人体行为特征的深度学习有效性。至此，关节行为模型构建完毕，一个行为检测方法基本形成，其框架结构如图 3。其中，位移矢量、骨骼夹角信息可用式 (1) 表达、计算，式中， V_t^m 表示第 t 帧中第 m 个关节的位移矢量， (x_{t-1}^m, y_{t-1}^m) 和 (x_t^m, y_t^m) 分别为 $t-1$ 帧、 t 帧中第 m 个关节的坐标， $\Delta\theta$ 为相邻两帧的骨骼夹角差，即骨骼夹角变化量。而 θ_{t-1}^m 和 θ_t^m 分别表示第 $t-1$ 帧、 t 帧中关节 m 的相邻两支骨骼夹角，其方法如式 (2)。

$$\begin{cases} V_t^m = \sqrt{(x_t^m - x_{t-1}^m)^2 + (y_t^m - y_{t-1}^m)^2} \\ \Delta\theta = \theta_t^m - \theta_{t-1}^m \end{cases} \quad (1)$$

$$\theta_t^m = \arccos \frac{a \cdot b}{|a| |b|}, |a| \neq 0, |b| \neq 0 \quad (2)$$

最后，位移矢量模型的输出为一个 1×14 的一维序列，可表示为 $(V_t^1, V_t^2, \dots, V_t^{13}, V_t^{14})$ 的形式，骨骼夹角变化模型的输出也是一个 1×14 的一维序列，可表示为 $(\Delta\theta_t^1, \Delta\theta_t^2, \dots, \Delta\theta_t^{13}, \Delta\theta_t^{14})$ 的形式。

为了印证位移矢量和骨骼夹角变化量表征人体行为信息的可行性，以 Li 等人^[20]于 2010 年建立的 MSR Action 3D 数据集开展实验，它记录了 10 个被试者、20 种类型的人体行为，每个被试者执行每个行为 2~3 次，总共包含了 567 个行为的关节运动序列。以挥手、侧踢的序列数据为例，计算每个行为的关节位移和骨骼夹角变化，并以关节半径的大小来直观表达该点与当前行为的相关程度，关节半径越大，则说明该点与当前行为的相关程度越高，反之则相关程度越低，结果如图 4。从图 4 可以看出人体在挥手和踢腿时，仅有手臂和腿部关节的半径明显增大，即手臂和腿

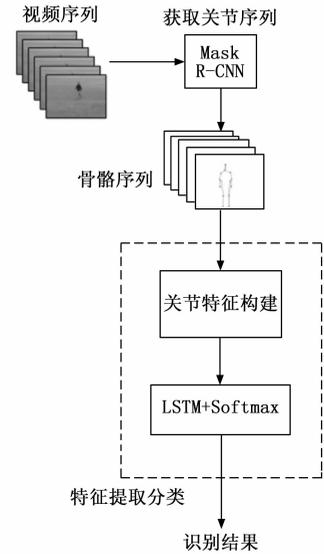


图 3 行为检测方法结构图

部的关节与当前行为具有较高的相关性，其余关节相关性不明显。说明补充这类信息可进一步明确行为识别应重点学习的关键部位。

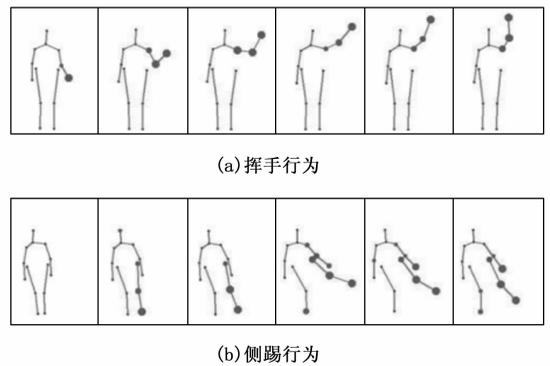


图 4 各关节的行为相关性表征图

3.3 行为特征提取

在通过关节行为模型计算得出位移矢量序列和骨骼夹角变化序列后，与原始关节点序列进行组合，形成一个尺寸为 4×14 的二维数组，以此作为前帧中人体的行为信息。考虑到人体行为具有连贯性和时间上的持续性，多帧行为信息序列相比于单帧行为信息能够更加完整表征人体行为，所以，对多帧行为信息进行特征提取的方式更加合理。

由于长时记忆网络 LSTM 在处理序列长时间依赖关系上展现出较好的效果，因此可以借助 LSTM 进行动作序列依赖特征提取并结合 Softmax 分类得出行为检测结果。LSTM 结构如图 5 所示。

LSTM 结构包括：信息输入门、输出门、控制状态信息是否保留的遗忘门和单元状态。LSTM 单元执行计算过程如式 (3) ~ (8)。输入门将输入信息与上一时刻的隐藏状态信息共同输入 sigmoid 函数以此来调节输出值控制本层的信息输入，式 (3) ~ (4) 为输入门计算公式，式 (3)

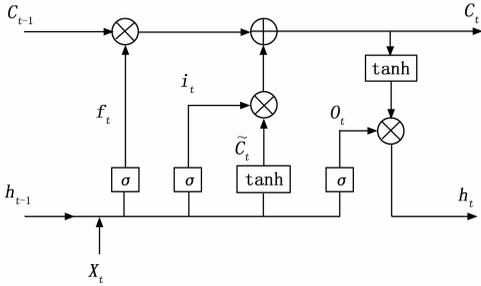


图 5 LSTM 结构图

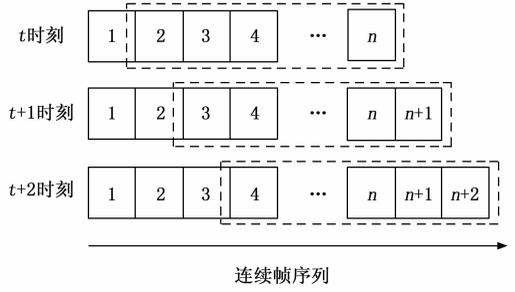


图 6 滑窗结构示意图

中，通过上一时刻的隐藏层输出 h_{t-1} 和 t 时刻的输入信息 x_t 计算 t 时刻的输入信息 i_t 。式(4)通过 h_{t-1} 和 x_t 计算 t 时刻的单元状态 \tilde{C}_t 。

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$\tilde{C}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

遗忘门机制能够将存储单元中的部分信息剔除，选择性地遗忘时间窗之前的信息。式(5)为遗忘门计算公式。

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

单元状态更新可用式(6)计算，在输入门和遗忘门的控制下结合上一时刻的单元状态信息来更新 t 时刻的单元状态。

$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t \tag{6}$$

式(9)、式(10)为 t 时刻隐藏层输出信息 h_t 。

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t * \tanh(C_t) \tag{8}$$

其中： w 为权重； b 为偏置； σ 是 sigmoid 激活函数； C_{t-1} 是 $t-1$ 时刻的单元状态； h_{t-1} 是 $t-1$ 时刻的隐藏层输出； x_t 是时间 t 的输入。

最终，设计的 LSTM 网络结构为两层，输入层送入行为序列信息，两个 LSTM 层进行时间序列特征提取，然后在第二个 LSTM 层后连接 Softmax 层对 LSTM 隐藏层的输出向量分类，得出行为分类结果。

另外考虑到检测系统的实时性要求以及多帧序列读取的问题，在多帧行为序列输入检测分类模块之前，还需设计一个特定帧长的滑动窗，将长视频的行为序列通过滑窗截取后再输入检测分类网络，图 6 为滑窗示意图。滑窗从首帧开始沿时间序列向后逐帧滑动，滑动过程中每检测到最新一帧的行为序列，就将新帧的行为序列加入到滑窗序列尾部，并删除滑窗序列头部的一帧行为序列，以此方式实现检测分类网络对长视频序列的固定帧数读取。

为了得到滑窗帧数长短的参考值，在 MSR Action 3D 数据集上展开分析、研究，该数据集记录了 10 个被试者、20 种类型的人体行为，每个被试者执行每个行为 2~3 次，总共包含了 567 个行为的关节运动序列。将每种行为的手、肘、膝、足部的横纵坐标沿时间序列绘制成关节运动时序曲线，从中统计每种行为的时序变化周期，指导滑窗帧长的确定，最终统计得出 20 种行为的平均周期帧长为 14.95，因此确定滑窗帧数长度取 15 帧。

4 实验结果与分析

系统设计完成后，首先需要在计算机上对行为检测网络进行训练，使其能够达到理想的检测效果，然后，将训练好的行为检测器部署到主控模块中运行并验证系统行为检测效果。

4.1 行为检测器训练与测试

为了能达到理想的行为识别效果，采用 MSR Action 3D 数据集对行为检测网络模型进行训练和测试。数据集集中的 20 类动作见表 2。从这 20 类动作中筛选出日常生活中最常见的高挥臂、水平挥臂、双臂挥动、捶打、前冲、拍手、侧拳、弯腰、前踢、侧踢、慢跑、接抛等 12 种行为对行为检测网络模型训练、测试。

表 2 MSRAction3D 数据集行为类别

1~10	11~20
高挥臂	双臂挥动
水平挥臂	侧拳
捶打	弯腰
手抓	前踢
前冲	侧踢
高抛	慢跑
画叉	网球发球
画勾	网球挥拍
画圆	高尔夫挥杆
拍手	接抛

实验训练平台基于 Windows10 操作系统，计算机处理器为英特尔酷睿 i5-10600kf，运行内存为 16G，显卡为 NVIDIA GTX1060，实验框架采用 TensorFlow 深度学习框架。

实验将 MSR Action 3D 数据集以 8:2 的比例划分为训练集和测试集，采用梯度下降算法对行为检测器进行迭代训练。为了能够充分训练从而获取性能良好的行为检测器，同时为避免行为检测器在训练过程中发生出现过拟合现象，使用早停法控制训练效果达到最佳时停止训练。当迭代训练次数达到 1 200 次时，识别准确率逐渐趋于稳定，训练停止。最终得出行为检测网络模型的识别准确率为 93.7%。

4.2 行为检测系统识别效果

将训练后得到的行为检测器部署到主控模块中，运行

该检测系统, 并对该系统的工作情况进行检验, 包括图像采集模块是否正常工作、显示模块是否正常显示画面和行为检测结果、报警模块能否针对危险行为类别发出警报。

经检验, 该系统各部分均能够正常工作, 摄像头采集画面正常、行为检测结果准确且能够在显示屏上正常显示, 当行为检测结果为危险行为时, 报警模块能够正常发出警报, 当行为检测结果不是危险行为时, 报警模块不工作, 屏幕仅显示当前行为类别, 测试达到预期效果, 系统实际工作情况如图 7。图 7 (a) 中目标人体执行挥手行为, 显示模块左侧显示真实画面, 右侧显示人体关节姿态, 同时屏幕显示当前行为类别为挥手, 报警模块二极管呈熄灭状态, 蜂鸣器不发出报警; 图 7 (b) 中目标进行出拳击打, 屏幕显示侧拳行为, 该行为被报警模块判断为危险行为, 则二极管闪烁提示, 蜂鸣器发出报警。

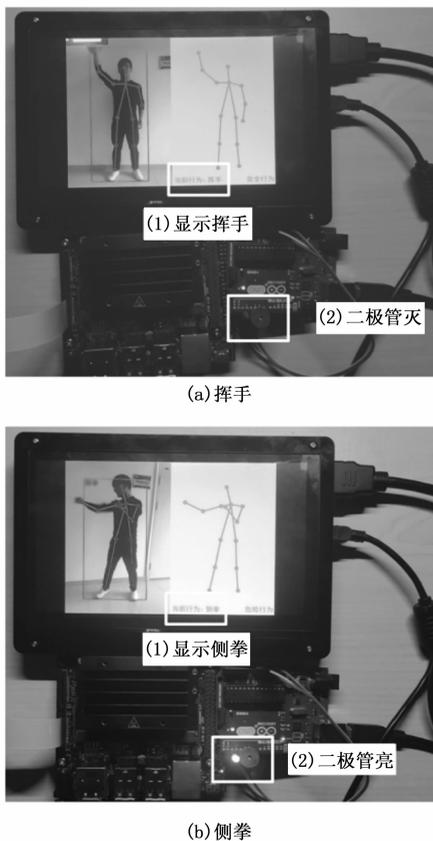


图 7 行为检测系统工作情况

5 结束语

针对现有行为检测系统中依赖惯性传感器、检测结果不够准确的问题, 设计了一种基于人体骨架信息的行为检测系统。以 Jetson nano 人工智能计算设备、图像采集模块和报警模块搭建硬件平台, 并以关节坐标为基础, 结合肢体间的运动差异, 建立了关节位移矢量和骨骼夹角变化的行为表征模型。在主控模块部署训练好的行为检测器, 提取行为表征模型中的人体行为特征并分类, 将分类结果输入报警模块进行危险行为判断并决策是否报警。实验结果

表明系统能够有效检测常见的人体行为。但对于复杂度较高和相似度较小的行为识别问题上仍存在不足, 今后将针对此类问题继续研究。

参考文献:

- [1] 马钰锡, 谭 励, 董 旭, 等. 面向智能监控的行为识别 [J]. 中国图象图形学报, 2019, 24 (2): 282-290.
- [2] 黄君浩, 贺 辉. 基于 LSTM 的眼动行为识别及人机交互应用 [J]. 计算机系统应用, 2020, 29 (3): 206-212.
- [3] 杜启亮, 黄理广, 田联房, 等. 基于视频监控的手扶电梯乘客异常行为识别 [J]. 华南理工大学学报 (自然科学版), 2020, 48 (8): 10-21.
- [4] 马敬奇, 雷 欢, 陈敏翼. 基于 AlphaPose 优化模型的老人跌倒行为检测算法 [J]. 计算机应用, 2022, 42 (1): 294-301.
- [5] 高莲弟, 陶 帅, 季长清, 等. 基于屋顶二值红外传感器网络的人体定位和行为识别系统设计 [J]. 计算机测量与控制, 2017, 25 (1): 163-166.
- [6] 王召军, 许志猛, 陈良琴. 基于红外阵列传感器的人体行为识别系统研究 [J]. 红外技术, 2020, 42 (3): 231-237.
- [7] 李鹏欢. 基于加速度传感器的自适应行为识别系统开发与实现 [D]. 杭州: 浙江工业大学, 2018.
- [8] 黄 龙. 基于惯性传感器的人体行为识别系统的研究与实现 [D]. 重庆: 重庆邮电大学, 2018.
- [9] 石磊鑫. 基于可穿戴设备的人体行为识别系统的研究 [D]. 青岛: 山东大学, 2020.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E, et al. ImageNet Classification with Deep Convolutional Neural Networks [J]. Neural Information Processing Systems. 2012, 141 (5): 1097-1105.
- [11] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos [J]. Advances in neural information processing systems, 2014, 27: 568-576.
- [12] ZHU Y, LAN Z, NEWSAM S, et al. Hidden two-stream convolutional networks for action recognition [C] // Asian conference on computer vision. Springer, Cham, 2018: 363-378.
- [13] ZHAO Y, MAN K L, SMITH J, et al. Improved two-stream model for human action recognition [J]. EURASIP Journal on Image and Video Processing, 2020, 2020 (1): 1-9.
- [14] JI S, XU W, YANG M, et al. 3D convolutional neural networks for human action recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013, 35 (1): 221-231.
- [15] ARUNNEHRU J, CHAMUNDEESWARI G, BHARATHI S P. Human action recognition using 3D convolutional neural networks with 3D motion cuboids in surveillance videos [J]. Procedia computer science, 2018, 133: 471-477.
- [16] CHEN J, XU Y, ZHANG C, et al. An improved two-stream 3D convolutional neural network for human action recognition [C] // 2019 25th International Conference on Automation and Computing (ICAC). IEEE, 2019: 1-6.

(下转第 44 页)