

基于卷积神经网络的足跟着地事件检测算法

李卓容¹, 王凯旋¹, 何欣龙¹, 糜忠良², 唐云祁^{1*}

¹中国人民公安大学刑事科学技术学院, 北京 100038;

²上海市现场物证重点实验室, 上海 200083

摘要 为解决基于可穿戴传感器的步态事件检测技术对个体配合程度依赖性大、能耗高、应用条件苛刻等问题, 提出一种基于机器视觉的足跟着地事件检测算法, 可以在不需要参与者合作的情况下, 利用普通摄像机实现对足跟着地事件的精确检测。提出一种新颖的特征, 即连续轮廓帧差图(CSD-maps)来表达步态模式。一个连续轮廓帧差图可以将视频帧中行人连续的轮廓二值图编码到一张特征图中, 使其蕴含丰富的步态时空信息。不同数量的行人连续轮廓帧差会产生不同的连续轮廓帧差图。利用卷积神经网络对连续轮廓帧差图进行特征提取和足跟着地事件分类。在公开数据库上, 对 124 名受试者在 5 个视角下不同穿着状态的视频数据进行训练和测试, 实验结果表明, 该方法具有良好的检测精度, 识别准确率达 93% 以上。

关键词 机器视觉; 步态事件检测; 连续轮廓帧差图; 卷积神经网络

中图分类号 TP29

文献标识码 A

doi: 10.3788/LOP56.211503

Heel-Strike Event Detection Algorithm Based on Convolutional Neural Networks

Li Zhuorong¹, Wang Kaixuan¹, He Xinlong¹, Mi Zhongliang², Tang Yunqi^{1*}

¹ School of Forensic Science, People's Public Security University of China, Beijing 100038, China;

² Shanghai Key Laboratory of Crime Scene Evidence, Shanghai 200083, China

Abstract In this study, we propose an algorithm based on machine vision to detect heel-strike events for solving the problem that the gait recognition technology based on wearable sensors is considerably dependent on the cooperation of participants, with high energy consumption and harsh application conditions. The proposed algorithm can accurately detect heel-strike events using ordinary cameras without the cooperation of participants. Initially, we develop an innovative feature for representing gait patterns by designing a consecutive-silhouette difference map (CSD-map). A CSD-map can encode the binary image of several consecutive pedestrian contours extracted from the video frames and output the combination as a single feature map. Different numbers of consecutive pedestrian contour differences result in different types of CSD-map. Further, a convolutional neural network is used for feature extraction and classification of the imaged heel-strike events. In a public database of video data for training and testing, we find 124 individuals under five angles in different wear conditions, and the experimental results obtained using these images denote the accuracy of our method. The identification accuracy is observed to be greater than 93%.

Key words machine vision; gait event detection; consecutive-silhouette difference maps; convolutional neural network

OCIS codes 150.1135; 330.5000; 070.5010; 100.4996

收稿日期: 2019-04-01; 修回日期: 2019-04-19; 录用日期: 2019-05-06

基金项目: 国家重点研发计划(2017YFC0822003)、国家自然科学基金(61503387, 61772539)、上海市现场物证重点实验室开放课题基金

* E-mail: tangyunqi@ppsuc.edu.cn

1 引言

步态是指走路时所表现的运动姿态,是一个连续的周期性运动。受运动习惯、伤残、疾病等因素的影响,不同的人的步态模式一般互不相同。对步态模式的研究和应用分为步态识别和步态分析两个方向。步态识别是模式识别领域的研究人员采用步态模式识别行人的身份;步态分析是医学领域的研究者将步态模式用于疾病诊断。这两种应用都是以步态周期事件检测为基础,因此步态周期事件的自动检测是步态识别和医学异常步态分析等人工智能应用的基础性问题^[1]。

通常,一个步态周期是从足跟着地到同侧腿足跟再次着地所经历的时间。一个周期可划分为多个独立事件^[2],根据腿的摆动特点,步态周期可分为支撑阶段和摆动阶段。Rose等^[3]基于运动学原理并根据标志性事件将步态进行了更详细的划分。对步态事件的准确检测可提高步态识别和分析的准确性。

目前步态事件检测方法主要分为两类:基于可穿戴传感器的步态事件检测方法和基于视觉的步态事件检测方法^[4]。基于可穿戴传感器的步态事件检测方法可以通过采集人体下肢关节和节段的运动数据准确检测步态事件。由于其精度较高,已经广泛应用于医学领域的异常步态评估。测力板和光学运动捕捉系统是最常见的步态检测设备,但价格昂贵、应用条件苛刻。另外,人们也使用一些可移动的、轻便的、成本低的检测设备,如嵌入在鞋底的足底压力开关、可穿戴的加速度计及陀螺仪等用于步态事件的检测^[5-9],但多数需要被检测人的高度配合。基于视觉的步态事件检测方法可以直接从单个或多个摄像机捕获的视频数据中检测步态事件,不需要任何其他特殊传感器的配合。与可穿戴传感器相比,摄像头更

便宜,使用也更方便。但是,光照、透视和服装的变化使得从二维视频数据中检测步态事件更具有挑战性,基于边缘、梯度等手工设计特征的判别性不强,对光照、透视等较为敏感。因此首先需要设计一个判别性较强的描述特征,其次需要选用合适的分类器。近年来研究中多用几何描述特征^[10-13],使用较多的分类器是支持向量机(SVM)^[14]和神经网络^[15-16]。

无论是在步态识别还是在步态分析的应用研究中,对步态周期的检测都依赖于对步态事件的检测。步态识别在步态轮廓图的基础上检测步态事件,对步态周期进行更清晰的划分,从而更好地识别受试者。步态分析建立在步态事件检测的基础上,能够对步态周期进行更精准的把握,从而对疾病及步态所反映的问题进行精确分析。为降低错误的人工标签对精度所产生的影响,本文选用较易识别的足跟着地事件进行步态事件检测。

本文提出一种基于二维视觉摄像头的足跟着地事件检测方法,利用视频数据来检测步态事件。设计了一种新颖的融合步态时空信息的步态表达特征,即连续轮廓帧差图(CSD-maps),将视频帧中行人连续的轮廓二值图,编码为一幅连续轮廓帧差图。针对CSD-maps特征图,基于卷积神经网络(CNN)构建足跟着地事件检测器。该方法可用于从不同视角下的视频数据中自动检测步态事件。

2 步态周期足跟事件检测方法

2.1 连续轮廓帧差图

通常,一个步态周期包括多个步态事件,主要分为足跟着地、足尖离地、足跟离地、足前支撑、足尖离地、双足对线(摆动中期开始)、胫骨垂直(减速摆动开始),如图1所示,图中数字表征对应阶段相对于一个完整步态周期的完成度。

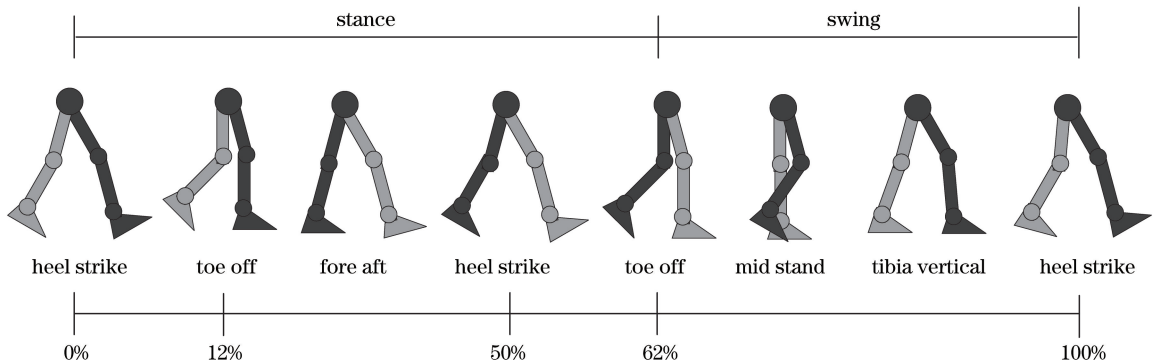


图1 一个步态周期内的步态事件

Fig. 1 Gait events within a gait cycle

对时空信息进行挖掘和融合是当前机器视觉领域研究的一个热点,而监控视频中恰好包含丰富的时空信息,如果能将其挖掘并利用,将会产生许多有意义的应用成果。当前行人检测与轮廓分割技术已经比较成熟,经典的方法有基于运动检测的算法^[17-19]、基于机器学习的方法^[20]、基于深度学习的方法^[21-22]等。本研究专注于步态事件检测,基于已有行人轮廓设计了步态事件检测算法,并加以实验验证。以 2-CSD 图为例说明连续轮廓帧差图的基本原理。

任意一幅 2-CSD 帧差图 F_i ,都是由视频序列的两个连续帧 I_{i-1} 和 I_i 生成的。对于 F_i 中的任意像素 $P_{2(j,k)}$,其像素值可以表示为

$$\text{value}(P_{2(j,k)}) = \begin{cases} 1, (P_{2(j,k)} \notin \Omega_{i-1}) \cap (P_{j,k} \in \Omega_i) \\ 2, (P_{2(j,k)} \in \Omega_{i-1}) \cap (P_{j,k} \notin \Omega_i), \\ 3, (P_{2(j,k)} \in \Omega_{i-1}) \cap (P_{j,k} \in \Omega_i) \end{cases} \quad (1)$$

式中, i 为当前帧数, j, k 分别为当前像素的行和列数, $P_{2(j,k)}$ 为 2-CSD 帧差图中任意一个像素, Ω_{i-1} 表示的是 I_{i-1} 帧像素集轮廓区域, Ω_i 代表 I_i 帧像素集轮廓区域。在实际应用中,行人轮廓是一种二值图像。两个连续轮廓生成的 2-CSD 图可以通过以下三个步骤计算。

1) 将像素灰度值从 I_{i-1} 复制到 F_i 。计算矩阵 I 为

$$I = I_i - I_{i-1}. \quad (2)$$

2) 根据矩阵 I 的值修改 F_i 的像素值,即

$$F_i(j,k) = \begin{cases} 1, \text{if } I(j,k) > 0 \\ 2, \text{if } I(j,k) < 0 \end{cases}, \quad (3)$$

式中, $F_i(j,k)$ 为 F_i 的像素值大小, $I(j,k)$ 为矩阵 I 第 j 行第 k 列值。

3) 对 F_i 像素值进行如下修改:

$$F_i(j,k) = \begin{cases} 3, \text{if } F_i(j,k) = 255 \\ F_i(j,k), \text{else} \end{cases}. \quad (4)$$

部分 2-CSD 图样本如图 2 所示。可以看出,与原始轮廓图像相比,2-CSD 图是足跟着地事件检测的显著特征。

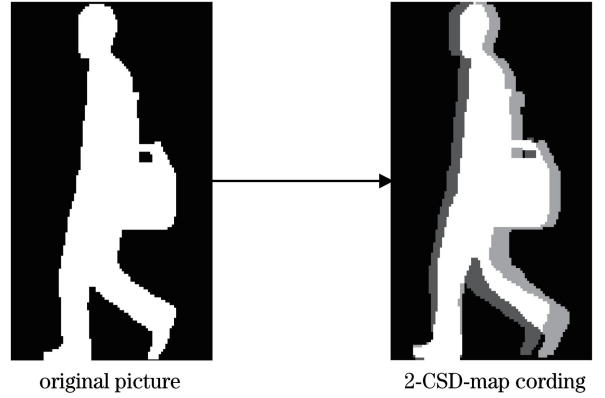


图 2 2-CSD 图样例

Fig. 2 Sample of 2-CSD-map

对于 n -CSD 图作以下定义:假设有 n 个连续的轮廓图像 I_1, I_2, \dots, I_n , 则有

$$\text{value}(P_{j,k}^n) = \begin{cases} 1, (P_{j,k}^n \in \Omega_1) \cap (P_{j,k} \notin \Omega_2) \cap (P_{j,k} \notin \Omega_3) \cap \dots \cap (P_{j,k} \notin \Omega_n) \\ 2, (P_{j,k}^n \notin \Omega_1) \cap (P_{j,k} \in \Omega_2) \cap (P_{j,k} \notin \Omega_3) \cap \dots \cap (P_{j,k} \notin \Omega_n), \\ 3, (P_{j,k}^n \notin \Omega_1) \cap (P_{j,k} \notin \Omega_2) \cap (P_{j,k} \in \Omega_3) \cap \dots \cap (P_{j,k} \notin \Omega_n) \end{cases} \quad (5)$$

式中 $\text{value}(P_{j,k}^n)$ 表示生成的 n -CSD-maps 中像素 $P_{j,k}^n$ 的像素值。给定 n 个连续轮廓帧差图,提取算法包括生成 CSD-maps 算法(算法 1,见图 3)和连续轮廓帧差图归一化算法(算法 2,见图 4)。本研究使用算法 2 将 CSD-maps 图像归一化到一定大小(如 $90 \text{ pixel} \times 140 \text{ pixel}$)。

2.2 卷积神经网络

卷积神经网络^[23]近年来发展迅速,并引起广泛重视。该网络不需对图像进行复杂的前期预处理,可以直接输入原始图像,在大型图像处理方面表现出色;且具有非常强的自主学习能力和高度非线性映射的能力,能够无监督学习非线性度量函数以解决跨视角步态识别问题。

卷积神经网络是一种具有多个互联层的前反馈

型深度学习网络,一个卷积神经网络主要包括卷积层、池化层、全连接层和 softmax 层。近年来,卷积神经网络在视觉识别领域取得了许多进展,包括图像分类^[24]、目标检测^[25]、场景解析^[26]等。在前反馈型网络大多效果不佳的时候卷积神经网络取得了出乎意料的效果,并且它突破了传统方法在图像分类方面的许多难关,如 ImageNet^[24]和人脸识别及数字识别等^[27]。与传统的基于特征的方法相比,卷积神经网络能够在不需要太多干预的情况下,通过反向传播算法学习特征并获得更高的精度。本研究选用经典卷积神经网络对步态周期中足跟着地事件进行预测,流程框架如图 5 所示,包括 3 个卷积层和 3 个全连接层。前 3 个卷积层分别有 64、128 和 256 个内核,大小分别为 5×5 、 $3 \times 3 \times 64$ 和 $3 \times 3 \times 128$ 。

Algorithm 1 Generate CSD-Maps

Input:

Continuous contour image: $I[w, h, n]$

The parameters w and h represent the width and height of the continuous contour image, respectively, and the parameter n represents the number of continuous contour images.

Output:

The CSD-map: Map

Code:

```

for  $i = 1$  to  $w$  do
  for  $j = 1$  to  $h$  do
     $t = I(i, j, :)$ ;
    value = 0;
    for  $k = 1$  to  $n$  do
      value = value +  $2^{(k-1)} * t(k)$ ;
    end for
    Map( $i, j$ ) = value;
  end for
end for
return Map;
    
```

图3 生成 CSD-maps 算法

Fig. 3 Algorithm for generating CSD-maps

Algorithm 2 Normalization for Consecutive Silhouette Difference Maps

Input:

1. Original CSD-Map image: OM
2. Normalize the width of the CSD-Map w
3. Normalize the height of the CSD-Map h

Output:

Standardized CSD-Map: NM

Code:

```

[x, y]=find(OM>0);
segm=OM(min(x):max(x), min(y):max(y));
NM=imresize(segm, [h,w]);
return NM;
    
```

图4 CSD-maps 归一化算法

Fig. 4 Algorithm for normalization of CSD-maps

第1个全连接层有1024个神经元,第2个全连接层有512个神经元,最后1个全连接层仅有2个神经元,一个用于识别足跟着地事件,另一个用于识别非足跟着地事件。最大池化层的核的大小为2,步长为2。

3 实验结果与分析

3.1 实验数据

以 CASIA 步态数据库(Dataset B)^[28]为实验数据进行实验,评估本文方法的准确性。本数据库的资料来自11个不同视角(0°、18°、36°、54°、72°、90°、108°、126°、144°、172°和180°)下的124名受试者(93名男性和31名女性)。在每个观看角度下,每名受试者在正常、外套、包三种不同的服装条件下,共采集10个视频。实验结果中,“nm”代表正常服装状态,“cl”代表穿着外套,“bg”代表携带背包。实验中对54°、72°、90°、108°、126°(视角1,总共大约400000帧)数据进行训练和测试。在0°、18°、36°、144°、172°和180°(视角2)的视角下拍摄的数据,两个连续帧图像轮廓相差很小,因此这部分数据未参与实验。该方法不能处理视角2下的视频数据,但对于大多数视角所捕获的视频数据都是有效的,这对于视频侦查很有意义。

本文实验将1~90号受试者(共计90人)的视频数据作为训练集,91~124号受试者(共计34人)的视频数据作为测试集,测试集数据量占比为27.4%。

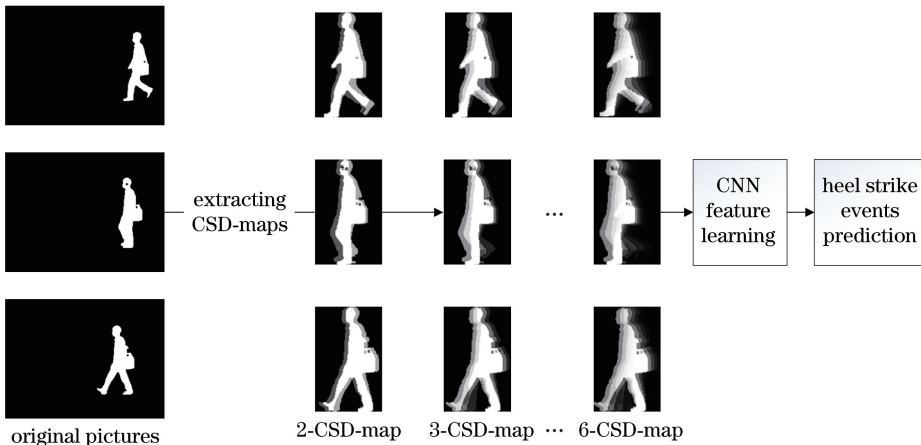


图5 流程框架

Fig. 5 Frame work of process

3.2 标签定义

实验前,所有的样本数据都需要手工标注标签值。因此,需首先明确足跟着地帧的定义。步

态是一种连续的、周期性的运动,医学领域将足跟着地事件定义为站立肢体接触地面的时刻。视频数据以一定帧率 θ 抽样记录人类步态。通常,一个

人的步态周期平均值为 1 s, θ 的经验值为 30 frame/s。所以记录一个人的一个步态运动周期大约需要 30 个连续帧, 间隔为 33 ms。但是问题在于 30 个连续的采样帧中, 可能不包括站立肢体接触地面的时刻。故本文将立肢着地后的帧定义为足跟着地帧。

特别地, 根据定义标记轮廓帧时可能会产生一定误差。如果足跟着地的时间恰好在 t_n 与 t_{n+1} 之间, 那么将第 $n+1$ 帧定义为足跟着地帧。当帧率为 θ 时, 两个连续帧之间的时间间隔为 $1/\theta$, 这意味着 $t_{n+1} - t_n = 1/\theta$ 。如果足跟着地事件发生在 (t_n, t_{n+1}) 周期中但靠近 t_n 时, 在 $n+1$ 帧时, 脚会在空中摇摆 $1/\theta$; 然而, 足跟着地事件发生在 (t_n, t_{n+1}) 周期中但靠近 t_{n+1} 时, 在 $n+1$ 帧时脚已经接触地面。后者误差并不影响该方法的有效性。

3.3 实验配置

实验使用 Caffe 框架^[29], Caffe 框架^[29] 是贾扬清博士在加州大学伯克利分校期间创建的一个深度学习框架, Caffe 通过逐层定义 (layer-by-layer) 的方式定义一个 Nets 网络, 该网络从数据输入层到损失层自下而上地定义整个模型。

n -CSD-maps 配置。针对不用规范化尺寸的 CSD-maps 图像进行实验, 结果表明, 不同尺寸的归一化 CSD-maps 对检测精度几乎没有影响。因此, 将 CSD-maps 的大小归一化为 $48 \text{ pixel} \times 32 \text{ pixel}$ 。 n -CSD-maps 的参数 n 设置为 2、3、4、5 和 6。

训练集和测试集的配置。选择每个视角的受试者 1 到受试者 90 的样本进行模型训练。其余样本 (受试者 91~124) 用于测试。

CNN 求解器的配置。初始学习率为 0.001, 动量为 0.9, 重量衰减为 0.0005。每个实验的最大迭代次数是 20000。CNN 中的权值初始化为零均值高斯分布初始化, 标准偏差为 0.01。

3.4 实验结果分析

该方法检测精度与 n -CSD 图的关系如图 6 所示。一般情况下, 随着 n 的增加, 检测的准确率略有提高, 原因是参数 n 越大, 在 CSD-maps 中编码的连续轮廓越多, 在 CSD-maps 中包含的信息也就越多。当参数 n 变大时, 检测精度大幅提高。例如, 在 72° 的视角下, 当参数 n 从 2 增加到 3 时, 该方法的准确率从 94.76% 增加到 95.02%。实验发现, 当编码的连续轮廓帧差图越来越多即 n 增大时, 检测精度也随之增大并在 $n=6$ 时精度最高。同时不难

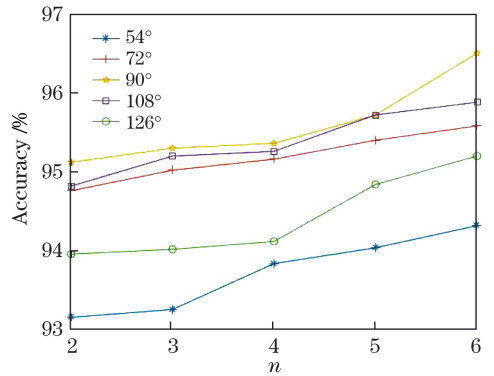


图 6 n -CSD-maps 检测精度折线图

发现, 90° 的检测精度最佳, 72° 与 108° 次之, 54° 与 126° 精度最差。

图 6 给出了该算法的检测精度, 实验结果表明本文算法的检测精度整体较好。由于检测精度随 n 增大并在 $n=6$ 时达到峰值, 因此使用各个视角下的 6-CSD 图绘制检测精度对比表。在 54° 的视角下, 利用 6-CSD 图, 该方法的准确率达到 94% 左右, 峰值达到 94.32%。在 72° 视角下, 利用 6-CSD 图, 该方法准确率达到 95% 左右, 峰值为 95.58%。在 90° 的视角下, 利用 6-CSD 图, 该方法准确率达到 96% 左右, 峰值为 96.50%。在 108° 视角下, 利用 6-CSD 图, 该方法准确率达到 95% 左右, 峰值为 95.88%。在 126° 视角下, 利用 6-CSD 图, 该方法的准确率也达到 94% 左右, 峰值达到 95.20%。

将本文算法与传统主成分分析 (PCA) 降维 SVM 算法在不同视角下的检测精度作对比。为了避免传统方法中 PCA 降维时产生高维运算, 首先将所有图片大小归一化至 48×32 , 根据得分判断当保留 95% 的数据时特征维度为 115 维, 将降维后图片维度归一化至 $[0, 1]$, 建立 SVM 分类器, 最后对足跟着地事件进行预测。通过交叉验证确定 SVM 分类器的最佳参数为 $c=2, g=1$ 。PCA+SVM 算法不同角度下检测精度折线图如图 7 所示, 图中精度均为 SVM 分类器在参数为 $c=2, g=1$ 时所得。实验表明, 随着 CSD-maps 编码图片的增多, PCA+SVM 算法精度整体呈上升趋势, 并在 6-CSD 时达到最高。由于 SVM 分类器并不能很好地通过支持向量在步态事件检测实验中找到最优的超平面, 进而将正负样本正确分类, 因此, 当提取特征较多时, 被错误预测的正负样本数量同时增多, 导致检测精度下降, 而当提取特征较少时, 预测出的正样本数量下降, 同样导致被错误预测的正负样本数量下降, 但

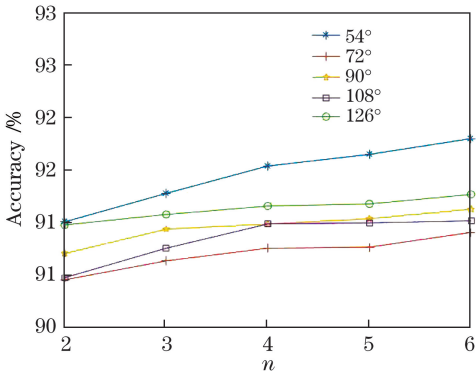


图7 PCA+SVM算法的 n -CSD-maps 检测精度折线图

Fig. 7 Line chart of detection accuracy of n -CSD-maps obtained by PCA+SVM algorithm

表现为检测精度的上升,即当偏角为 54° 和 126° 时,检测精度反而要高于偏角小时,但并不是真正意义上检测精度的提升。

由图6与图7可以看出,本文方法在侧面识别时即摄像机与行人行走路线夹角为 90° 时识别效果最好,而在夹角较大的情况下仍取得了较好识别效果。本文方法较 PCA+SVM 算法,准确率和识别效果整体均有较明显提升,能够较好解决 PCA 要求各主元之间关系是线性的且待处理数据必须服从指数型概率分布模型而导致识别率不高以及 SVM 分类器在本实验中很难找到最优超平面的难点问题。为了更直观对比两种算法的检测精度,绘

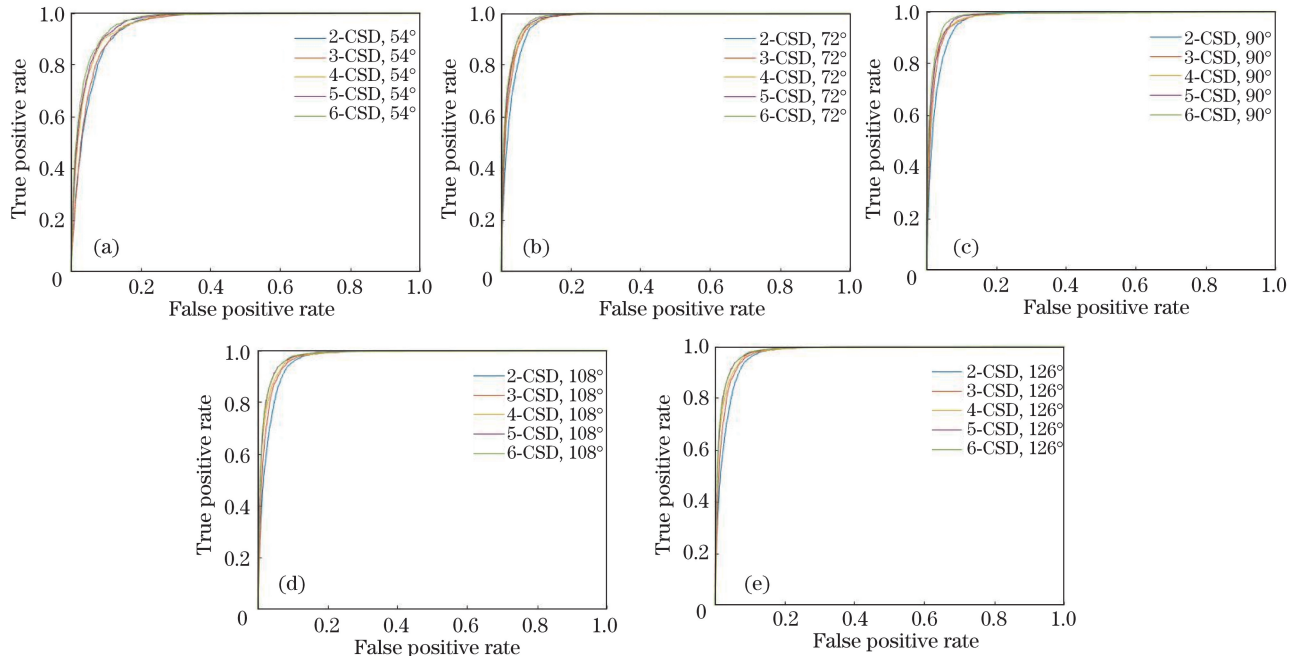


图9 不同视角的 ROC 曲线。(a) 54° ; (b) 72° ; (c) 90° ; (d) 108° ; (e) 126°

Fig. 9 ROC curves of different angles. (a) 54° ; (b) 72° ; (c) 90° ; (d) 108° ; (e) 126°

制对比柱状图,如图8所示。

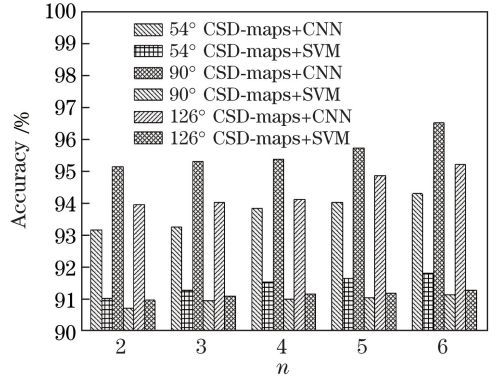


图8 CNN算法与 SVM 算法检测精度对比柱状图
Fig. 8 Histogram comparison of detection accuracy of CNN and SVM algorithms

采用 n -CSD-maps 的方法绘制在不同视角下的 ROC(Receiver Operating Characteristic) 曲线如图9所示,一般情况下,该方法在冠状面视角周围的检测精度要高于矢状面视角。特别是在 90° 视角下,该方法的准确率峰值达 96.50% ,明显高于其他视角。这表明,从矢状面视角捕获的视频数据生成的 CSD-maps 在步态事件检测中包含的有用信息少于冠状面视角,其原因是在矢状面视角下拍摄的视频帧的连续轮廓与冠状面视角下拍摄的视频帧的连续轮廓差异较小。

由以上实验结果可知,识别效果以 90° 为中心对称下降, 90° 时最优临界点最靠近1; 72° 和 108° 时

下降幅度不大,54°时下降幅度较大,126°时略有波动。总体识别效果良好。

为了测试交叉视角的检测精度,使用6-CSD图的混淆矩阵来评估该方法,结果如表1所示,峰值已加粗表示。不难看出,该方法在对角线上的准确率最高,在其他区域准确率在90%左右,结果表明在交叉视点的检测中准确率较高。同时,建立所有角度下的训练模型,其ROC曲线图如图10所示。实验结果表明,本文方法取得了很好的效果。

表1 利用6-CSD检测交叉视角精度的混淆矩阵

Table 1 Confusion matrix when 6-CSD is used to obtain detection accuracy of cross angle

Angle	54°	72°	90°	108°	126°
54°	0.9432	0.9292	0.9274	0.9288	0.9230
72°	0.9404	0.9558	0.9448	0.9432	0.9336
90°	0.9348	0.9372	0.9650	0.9496	0.9414
108°	0.9330	0.9252	0.9486	0.9588	0.9456
126°	0.9276	0.9104	0.9028	0.9268	0.9520

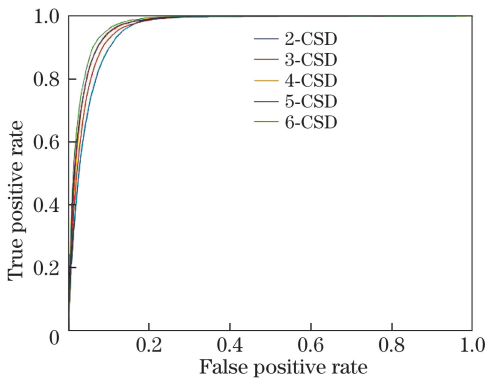


图10 所有视角下的ROC曲线

Fig. 10 ROC curves from all angles

4 结 论

提出一种基于卷积神经网络的足跟着地事件检测算法,该算法仅依靠视频数据识别步态事件,解决了基于穿戴式传感器的步态事件识别所需设备价格昂贵、应用条件苛刻、需要被测试者高度配合等瓶颈问题。本文算法一方面提出了一种新颖的连续轮廓帧差图的特征,使得足跟着地帧更易被识别,另一方面使用卷积神经网络进行特征降维并对足跟着地事件进行分类。结果表明传统PCA降维SVM分类算法分类正确率峰值不足92.5%,而本文算法分类正确率达93%以上,整体效果较优。下一步的研究重点将是在更大的数据库中测试算法,并进一步针对不同条件测试足跟着地事件分类的实际性能。

参 考 文 献

- [1] Muro-de-la-Herran A, Garcia-Zapirain B, Mendez-Zorrilla A. Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications[J]. *Sensors*, 2014, 14(2): 3362-3394.
- [2] Giroux M, Moissenet F, Dumas R. EMG-based validation of musculo-skeletal models for gait analysis [J]. *Computer Methods in Biomechanics and Biomedical Engineering*, 2013, 16 sup1: 152-154.
- [3] Rose J, Gamble J G. *Human walking*[M]. 2nd ed. Baltimore: Williams & Wilkins, 1994.
- [4] Yang C, Ugbohue U, Carse B, *et al.* Multiple marker tracking in a single-camera system for gait analysis[C] // 2013 IEEE International Conference on Image Processing, September 15-18, 2013, Melbourne, VIC, Australia. New York: IEEE, 2013: 3128-3131.
- [5] Huang B F, Chen M, Shi X, *et al.* Gait event detection with intelligent shoes [C] // 2007 International Conference on Information Acquisition, July 8-11, 2007, Seogwipo-si, Korea. New York: IEEE, 2007: 579-584.
- [6] Catalfamo P, Moser D, Ghousayni S, *et al.* Detection of gait events using an F-Scan in-shoe pressure measurement system[J]. *Gait & Posture*, 2008, 28(3): 420-426.
- [7] Heliot R, Pissard-Gibollet R, Espiau B, *et al.* Continuous identification of gait phase for robotics and rehabilitation using microsensors[C] // ICAR '05. Proceedings., 12th International Conference on Advanced Robotics, 2005, July 18-20, 2005, Seattle, WA, USA. New York: IEEE, 2005: 686-691.
- [8] Williamson R, Andrews B J. Gait event detection for FES using accelerometers and supervised machine learning [J]. *IEEE Transactions on Rehabilitation Engineering*, 2000, 8(3): 312-319.
- [9] Hanlon M, Anderson R. Real-time gait event detection using wearable sensors [J]. *Gait & Posture*, 2009, 30(4): 523-527.
- [10] Tang S, Wang X Y, Lü X T, *et al.* Histogram of oriented normal vectors for object recognition with a depth sensor[M] // Lee K M, Matsushita Y, Rehj J M, *et al.* European conference on computer vision-ACCV 2012. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2013, 7725: 525-538.

- [11] Huang C H. Research on algorithms of human action recognition based on videos[D]. Chengdu: University of Electric Science and Technology of China, 2016.
黄成挥. 基于视频的人体行为识别算法研究[D]. 成都: 电子科技大学, 2016.
- [12] Wang J. Human activity recognition using multiple instance learning[J]. *Information Technology*, 2016, 40(7): 65-70.
王军. 基于多示例学习法的人体行为识别[J]. *信息技术*, 2016, 40(7): 65-70.
- [13] Liu Z, Dong S D. Study of human action recognition by using skeleton motion information in depth video [J]. *Computer Applications and Software*, 2017, 34(2): 189-192, 219.
刘智, 董世都. 利用深度视频中的关节运动信息研究人体行为识别[J]. *计算机应用与软件*, 2017, 34(2): 189-192, 219.
- [14] Zhang Y J, Wang H M, Fu X H, *et al.* Identification of steel plate damage position based on particle swarm support vector machine[J]. *Chinese Journal of Lasers*, 2017, 44(10): 1006006.
张燕君, 王会敏, 付兴虎, 等. 基于粒子群支持向量机的钢板损伤位置识别[J]. *中国激光*, 2017, 44(10): 1006006.
- [15] Liu F, Shen T S, Ma X X. Convolutional neural network based multi-band ship target recognition with feature fusion[J]. *Acta Optica Sinica*, 2017, 37(10): 1015002.
刘峰, 沈同圣, 马新星. 特征融合的卷积神经网络多波段舰船目标识别[J]. *光学学报*, 2017, 37(10): 1015002.
- [16] Bi L H, Liu Y C. Plant leaf image recognition based on improved neural network algorithm[J]. *Laser & Optoelectronics Progress*, 2017, 54(12): 121102.
毕立恒, 刘云潺. 基于改进神经网络算法的植物叶片图像识别研究[J]. *激光与光电子学进展*, 2017, 54(12): 121102.
- [17] Barnich O, van Droogenbroeck M. ViBe: a universal background subtraction algorithm for video sequences [J]. *IEEE Transactions on Image Processing*, 2011, 20(6): 1709-1724.
- [18] Wang H Z, Suter D. A consensus-based method for tracking: modelling background scenario and foreground appearance [J]. *Pattern Recognition*, 2007, 40(3): 1091-1105.
- [19] Hofmann M, Tiefenbacher P, Rigoll G. Background segmentation with feedback: the pixel-based adaptive segmenter [C] // 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, June 16-21, 2012, Providence, RI, USA. New York: IEEE, 2012: 38-43.
- [20] Benenson R, Omran M, Hosang J, *et al.* Ten years of pedestrian detection, what have we learned?[M] // Agapito L, Bronstein M, Rother C. European conference on computer vision-ECCV 2014 Workshops. Lecture notes in computer science. Cham: Springer, 2015, 8926: 613-627.
- [21] Li J N, Liang X D, Shen S M, *et al.* Scale-aware fast R-CNN for pedestrian detection [J]. *IEEE Transactions on Multimedia*, 2018, 20(4): 985-996.
- [22] Zhang L L, Lin L, Liang X D, *et al.* Is faster R-CNN doing well for pedestrian detection?[M] // Leibe B, Matas J, Sebe N, *et al.* European conference on computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9906: 443-457.
- [23] LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision[C] // Proceedings of 2010 IEEE International Symposium on Circuits and Systems, May 30-June 2, 2010, Paris, France. New York: IEEE, 2010: 253-256.
- [24] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C] // Advances in neural information processing systems 25 (NIPS 2012), December 3-8, 2012, Harrahs and Harveys, Lake Tahoe. New York: NIPS, 2012: 1097-1105.
- [25] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 580-587.
- [26] Farabet C, Couprie C, Najman L, *et al.* Learning hierarchical features for scene labeling [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(8): 1915-1929.
- [27] Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering[C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 815-823.
- [28] Yu S Q, Tan D L, Tan T N. A framework for evaluating the effect of view angle, clothing and

carrying condition on gait recognition [C] // 18th International Conference on Pattern Recognition (ICPR '06), August 20-24, 2006, Hong Kong, China. New York: IEEE, 2006: 441-444.

[29] Jia Y Q, Shelhamer E, Donahue J, *et al.* Caffe:

convolutional architecture for fast feature embedding[C]// Proceedings of the 22nd ACM international conference on Multimedia, November 3-7, 2014, Orlando, Florida, USA. New York: ACM, 2014: 675-678.