

DOI:10.19651/j.cnki.emt.2208819

基于改进型对抗网络的步态特征提取方法研究*

李言¹ 曾维² 蒋毅¹ 王玥妲² 罗伟洋¹ 于真²

(1. 成都理工大学机电工程学院 成都 610051; 2. 成都理工大学计算机与网络安全学院(牛津布鲁克斯学院) 成都 610051)

摘要: 针对步态识别易受环境干扰等问题,本文以步态特征提取方法为研究重点,基于对抗学习网络框架提出了改进型姿态估计算法提取步态特征。该方法利用改进型残差网络获取由低层次到高层次的步态特征,随着网络层数的加深,对残差网络做出相应的调整,突出对局部细节特征信息的聚焦;同时设计了时序编码器,不仅提高了步态特征对于环境变化的泛化性,还减少了环境对特征提取的影响。最终在三种不同的实验模式下,基于CASIA数据集进行了大量的实验,识别精度均在83%以上,最终证明本文所提出的特征提取方法在复杂环境展现出良好的灵活性。

关键词: 特征提取;对抗网络;步态识别;姿态估计;预处理

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 510.20

Research on gait feature extraction method based on improved generative adversarial networks

Li Yan¹ Zeng Wei² Jiang Yi¹ Wang Yue-dayi² Luo Weiyang¹ Yu Zhen²

(1. School of Mechanical and Electrical Engineering, Chengdu University of Technology, Chengdu 610051, China;

2. School of Computer and Network Security(Oxford Brooks College), Chengdu University of Technology, Chengdu 610051, China)

Abstract: Aiming at the problem that gait recognition is vulnerable to environmental interference, this paper focuses on the gait feature extraction method, an improved attitude estimation algorithm is proposed based on the anti learning network framework to extract gait features. The improved residual network is used to obtain the gait features from low level to high level. With the deepening of the number of network layers, the residual network is adjusted accordingly to highlight the focus on the local detail feature information; A timing encoder is designed, which not only improves the generalization of gait features to environmental changes, but also reduces the impact of environment on feature extraction. Finally, a large number of experiments are carried out based on CASIA data set under three different experimental modes, the recognition accuracy is more than 83%, which finally proves that the feature extraction method proposed in this paper shows good flexibility in complex environment.

Keywords: feature extraction; generative adversarial networks; gait recognition; attitude estimation; pretreatment

0 引言

特征提取是计算机视觉领域的一项重要任务,其目标是从复杂场景中检测特定关键信息,输出全部或局部信息,有着较为广泛的应用前景。随着现代科技的发展,基于生物特征的识别技术(biometric identification technology, BIT)识别准确率逐渐提升,不同的识别方案被用于不同场景^[1]。Das等^[2](2015)提出了两阶段视角不变的多尺度步态识别方法(two-phase view-invariant gait recognition method, VI-MGR),该方法基于加权随机子空间学习进行分类,可避免子空间过度学习从而提高识别效果。邓秀春

等^[3](2011)提出了三维的人体骨骼点的步态识别系统,与二维的步态分析系统相比,其识别率得到了改进。Wu等^[4](2017)提出了基于(convolutional neural network, CNN)的步态识别,对于多视角的步态分析优于同时期方法。为了增强跨视图条件下步态识别的鲁棒性,Eifly等^[5](2017)提出了几何视图变换模型(geometry view transformation model, GVTM),该方法通过自由形式的变形框架将两个不同视图的步态特征从几何形状转换为中间视图的步态特征。该方法对特征的变换更加高效。为了解决子空间学习方法在相似度度量依赖数据的问题,Chen等^[6](2020)提出了新的无监督学习方法,该方法可以自适

收稿日期:2022-01-10

* 基金项目:国家重点研发项目(2018YFC1505102)资助

应地学习交叉视点样本之间的相似信息。胡叶等^[7](2021)提出基于 K-means 聚类算法的行人步态识别方法,对运动状态有较好的识别效果。

目前,对步态序列的处理方法主要有以下两种:按照静态图像帧的方式进行处理,这种方法虽然易于实现但是很容易丢失空间信息^[8];按照连续视频序列,即从原步态轮廓序列数据中直接提取步态特征^[9]。虽然人体行走的情况复杂多变,但是在时间属性上,行走过程是具有周期属性的^[10]。基于上述分析,本文将步态轮廓视为一组步态序列轮廓图,作为周期性运动,步态可以由一个周期表示。步态识别算法可以归为两类:第一种是基于模型的方法,跟踪分析模型的参数,利用参数进行识别;第二种是基于非模型方法,直接对步态视频序列进行分析而不需要预先假定的模型^[11],与基于模型的方法相比,非模型的方法能够直接从轮廓中提取步态描述信息^[12]。传统的应用于步态特征提取方法往往依靠梯度方向直方图(histogram of oriented gradient, HOG)和尺度不变特征变换(scale invariant feature transform, SIFT)^[13]等。虽在特定的场景达到了较高的检测速度与精度,但该方法依赖先验知识,导致自适应性及泛化性较差。本文提出了基于改进型姿态估计的步态特征提取。与传统步态特征提取方法不同,该方法可以减少复杂背景对特征提取的干扰。不仅提高了步态特征对于环境变化的泛化性,还在一定程度上减少了人体遮挡对步态特征提取的影响。同时,利用提出的周期步态原始特征图像作为识别特征集合的方法,能够保留更多的原始特征以提高步态识别率。

1 基于姿态估计的特征提取整体架构

本文提出的姿态估计的实现方法基于对抗学习网络框架^[14],首先通过该框架区分大型动作捕捉数据集(archive of motion capture as surface shapes, AMASS)里真实的人类动作和由深度学习网络线性回归生成的人体姿态;然后,通过对抗学习逐渐提高回归网络生成的人体姿态的准确性;最后,通过预处理将提取的步态特征生成步态特征集合。通过预处理可以有效去除环境噪声,突出步态特征,极大提升步态识别算法识别率。基于姿态估计的步态特征提取总体架构如图 1 所示。

2 特征提取的改进方法

现有人体姿态估计往往不能产生准确的预测,其中一个主要的原因是直接从 2D 的视频图像预测出的 3D 人体姿态真实性不足^[15-17]。本文所采用的姿态估计是基于对抗学习网络框架获得高质量的人体三维运动姿态特征。姿态估计总体流程如图 2 所示。

2.1 改进型残差网络

为了解决深度神经网络在图像识别上存在的问题^[18-19],并提升网络的效率,通过改进型网络利用残差连接

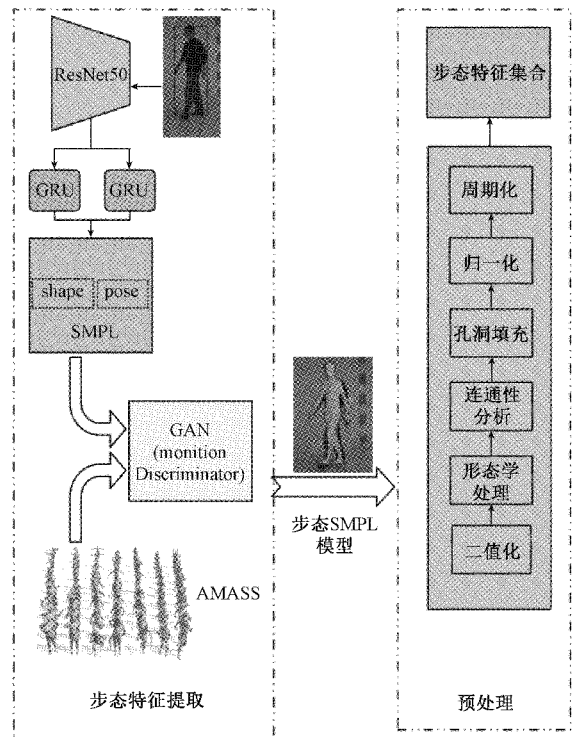


图 1 基于姿态估计的步态特征提取总体流程

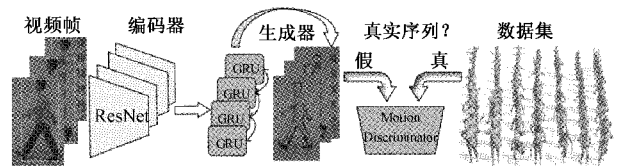


图 2 姿态估计总体流程

的方式代替 Highway 的网关单元。残差网络不仅可以提升网络训练的速度,还可以明显地提升网络的准确率。并且由于残差网络的结构,让更深层次的网络训练成为了可能。残差网络基本结构如图 3 所示。

2.2 时序编码器

考虑人体的运动不仅在空间上呈现连续,在时间上也连续,因此添加时序编码器。对于一个视频序列帧的集合,其前后视频帧的信息呈现连续性,具体表现为一个视频序列里,可以从过去的视频帧推断未来视频帧的信息特征。尤其是在人体姿态估计任务中,当目标人体的视频序列部分存在遮挡,无法直观地从源视频获取对应姿态时,可以从邻近无遮挡的图片序列推测人体肢体的运动。因此,过去的肢体特征信息可以帮助约束未来的肢体特征表达。

1) 时序编码器基本框架设计

设计的时序编码器包括 GRU 层, SMPL 线性回归层。该编码器以空间特征编码器输出的向量 $f_i \in \mathbb{R}^{2048}$ 作为输入,并且输出 SMPL 的 82 个参数序列。时序编码器的作用是对于输入的连续的图片序列 I_1, \dots, I_T , 输出每张图片人物对应的姿态和形状参数,并且将空间特征编码器输出

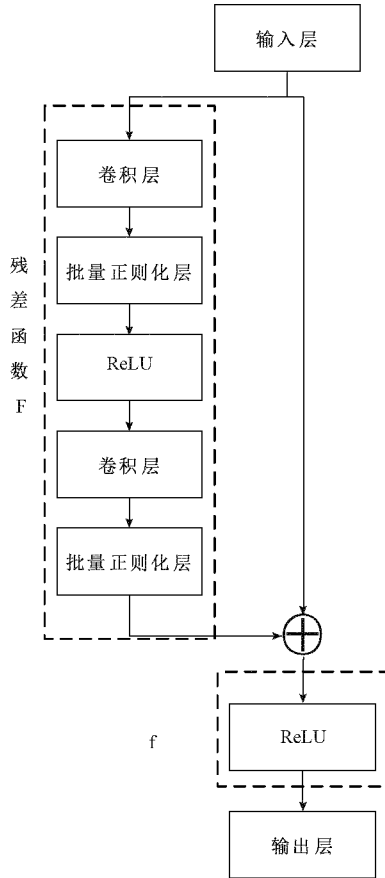


图 3 残差网络结构

的向量 $f_i \in \mathbb{R}^{2048}$ 作为时序编码器内部门控循环单元网络层的输入,并且根据之前的视频帧生成潜在的特征向量 g_i ,表达式如 $g(f_1), \dots, g(f_T)$,然后,使用 g_i 作为线性回归层的输入,同时将输入迭代反馈。线性回归器使用平均位姿作为初始化的参数,每次迭代的时候 g_i 作为输入。时序编码器网络框架如图 4 所示。

2) SMPL 线性回归

将经过时序编码器处理后的视频帧特征输入 SMPL 线性回归层,得到 82 个 SMPL 参数,获得人体姿态。SMPL 模型中,人体是一个形状不变的基本模型,并受到一定约束的形变总和。SMPL 包括形状参数 β 和姿态参数 θ 。其中,通过 PCA 得到人体的形状参数 $\beta^{[10]}$,其中 β 参数代表人体的高矮胖瘦,身体比例等参数。使用运动树表示人体的姿态,通过三维向量表示各个节点之间的旋转空间关系,最终关节的部分旋转向量构成了姿态参数 θ 。

SMPL 模型通过 θ 符表征一个人的形状和姿态,其中包括姿态参数 $\theta (\theta \in \mathbb{R}^{72})$ 以及形状参数 $\beta (\beta \in \mathbb{R}^{10})$,姿态参数包括身体整体旋转和 23 个关节轴角格式的相对旋转。形状参数为主成分形状空间的前 10 个系数。SMPL 输出 3D 的网格,基本计算公式如下:

$$M(\theta, \beta) \in \mathbb{R}^{6890 \times 3} \quad (1)$$

其中, M 是 SMPL 功能函数,生成 6 890 个顶点的人体

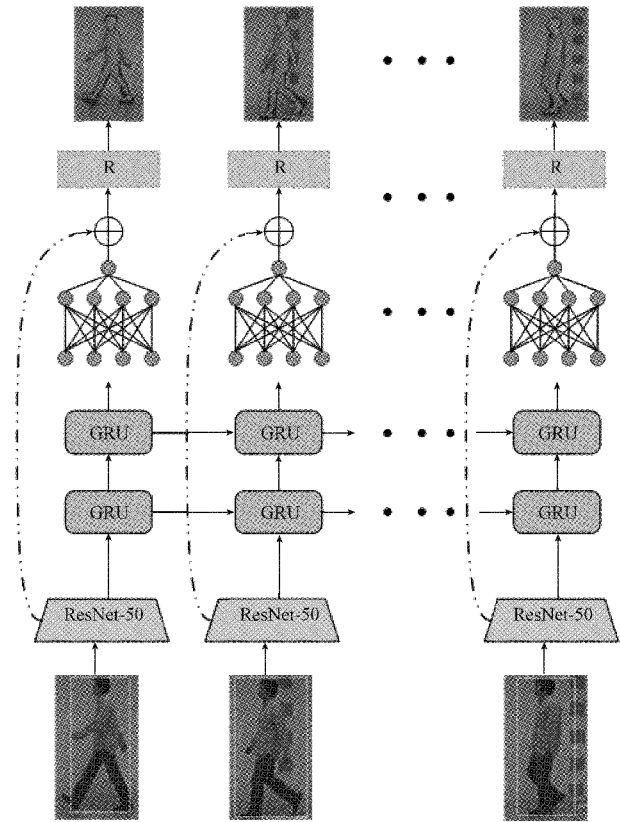


图 4 时序编码器网络框架

蒙皮模型,人体最终坐标计算方式如下:

$$M(\beta, \theta) = W(T_p(\beta, \theta), J(\beta), \theta, w) \quad (2)$$

其中, W 是蒙皮函数, LBS 是典型的蒙皮函数, LBS 函数中每个点由每个骨架的旋转矩阵的加权和表示,计算公式如下。

$$\bar{l}'_i = \sum_{k=1}^K w_k \cdot iG_k(\theta, J) \bar{t}_i \quad (3)$$

$G_k(\theta, J) \bar{t}_i$ 是由旋转角度 θ 和骨架关键点 J 决定的变换矩阵。 T_p 表示经由人体基本胖瘦参数确定以后的姿势变化。

$$T_p(\beta, \theta) = \bar{T} + B_s(\beta) + B_p(\theta) \quad (4)$$

$B_s(\beta)$ 是人体高矮胖瘦参数的模型位移, $B_p(\theta)$ 是即将发生的人体位移,方便后续步骤的旋转操作。

综上所述,一个变化后的人体顶点最终表示函数如下:

$$l'_i = \sum_{k=1}^K w_k \cdot iG'_k(\theta, J(\beta)) (\bar{l}_i + b_s, ((\beta) + b_p, (\theta))) \quad (5)$$

其中, $J(\beta)$ 是求相关骨骼关键点函数。给定一个视频序列,时序编码器中线性回归层的网络结构计算公式如下:

$$\theta = [(\hat{\theta}_1, \dots, \hat{\theta}_t), \hat{\beta}] \quad (6)$$

其中, $\hat{\theta}_t$ 是对应时间 t 的姿态参数, $\hat{\beta}$ 是序列中单人的

形状估计。首先对于每一帧图片预测单人的形状参数,然后在整体序列上平均得到 $\hat{\beta}$ 。通过时序编码器输出的 SMPL 人体模型与来自 AMASS 数据集的样本一起输入动作鉴别器 D_M ^[9],判断网络是否生成正确的人体姿态。

2.3 动作鉴别器

SMPL 线性回归生成的姿态是不确定的,为了排除无效的人体姿态,本文设计动作鉴别器 D_M 判断生成的人体姿态是否符合真实空间。

1) 动作鉴别器基本框架设计

在运动鉴别器中,输出 $\hat{\theta}$ 作为多层 GRU 模型的输入,并作为生成器,在每一个时间步骤 I ,预估一个隐藏的特征值, $h_i = f_m(\hat{\theta}_i)$,为了聚合隐藏的状态, $[h_i, \dots, h_T]$,使用了自注意力机制。最后,线性层输出预测值,范围 $\in [0, 1]$,预测 θ 代表的人体姿态属于真实人体序列的概率。

反向传播到生成器 G 的对抗损失公式如下:

$$L_{adv} = E_{\theta \sim PR} [(D_M(\hat{\theta}) - 1)^2] \quad (7)$$

$$L_{D_M} = E_{\theta \sim PR} [(D_M(\hat{\theta}) - 1)^2] + E_{\theta \sim PR} [(D_M(\hat{\theta}))^2] \quad (8)$$

其中, PR 是来自 AMASS 数据集的真实动作序列。 PR 是生成的动作序列。 D_M 动作鉴别器被真实数据训练,并且学习多种真实的动作姿态数据。动作鉴别器包括 GRU 和自注意力机制,网络框架如图 5 所示。

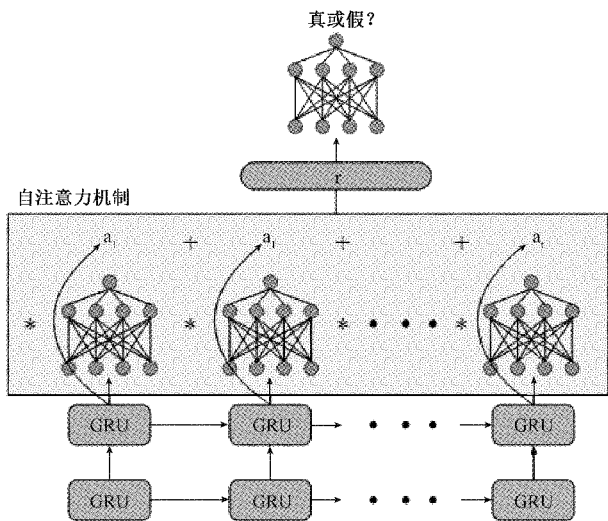


图 5 运动鉴别器

3 实验结果与分析

3.1 CASIA-B 数据集测试分析

为了验证提出的步态特征提取方法,源视频采用 CASIA 数据集的原始步态视频,该系列视频按照不同的视角,不同的行走状态对视频序列编号。其中视频序列的视角定义如图所示,共有 11 个视角,从 0° 到 180° ,视角间隔

18° ,如图 6 所示,图中列出了关键的几个视角,其中,视角的角度定义为摄像机与行人行进方向的夹角形成的角度,例如 0° 是和行人行进角度平行, 90° 是和行进角度垂直,其余的角度以此类推。

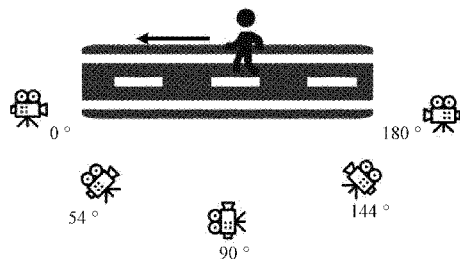


图 6 步态视频视角定义

1) 不同视角测试

按照不同的视角条件对步态视频进行特征提取,具体特征提取效果图如图 7 所示,图中展示了在不同视角下特征提取效果图。



图 7 不同视角特征提取效果图

2) 不同行走条件

如图 8 所示,在 90° 的视角的不同行走条件的特征提取效果图,(a)是在带包行走条件下的姿态估计,可以明显区分人体和携带的包裹;(b)是正常行走条件下的姿态估计,

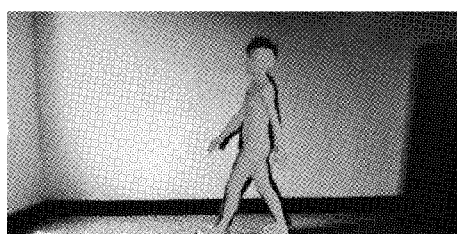


图 8 不同行走条件特征提取效果图

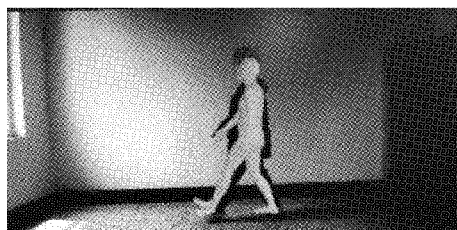
可以勾画出清晰的人体轮廓;(c)是大衣行走条件下的姿态估计,即使无法直观观测到行人躯体,姿态估计也可以勾画出隐藏在大衣下的躯干。因此,基于姿态估计的步态特征提取方法对环境具有更好的抗干扰性,并且可以在一定程度上解决人体遮挡问题。

3.2 实际测试分析

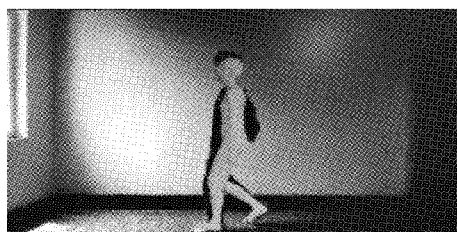
实际环境的测试是通过摄像头采集不同行走状态和不同视角的行走姿态验证得到。为了更好地验证,和CASIA-B数据集一样,主要包括三种行走条件,如图9所示。分别是正常,带包,大衣3种行走条件下的姿态估计。



(a) 正常行走条件步态测试



(b) 大衣行走条件步态测试

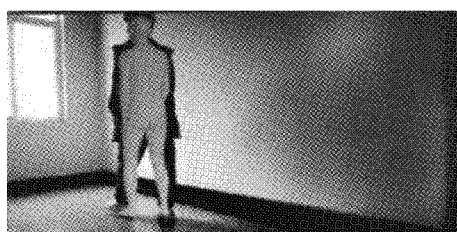


(c) 带包行走步态测试

图9 不同步态下的实际测试

按照视角的不同,可以分为 36° 视角, 72° 视角, 108° 视角, 144° 视角。如图10所示,在大衣行走条件下对步态测试4个视角的步态。

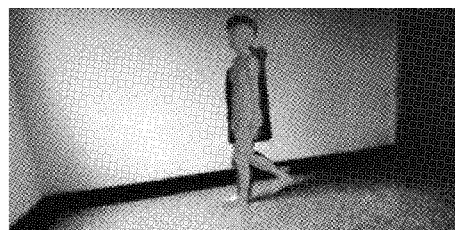
如表1所示,将不同行走条件的步态剪影图作为特征输入时,本文提出的特征提取方法在不同的行走条件下识别率均高于83%。



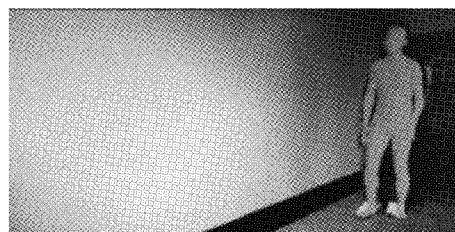
(a) 36° 的大衣行走实际环境步态



(b) 72° 的大衣行走实际环境步态



(c) 108° 的大衣行走实际环境步态



(d) 144° 的大衣行走实际环境步态

图10 不同视角下的大衣行走步态

表1 不同行走条件下的识别率

条件	准确率	条件	准确率
正常+背包	87.9	正常	89.5
正常+大衣	85.8	背包	84.2
背包+大衣	84.6	大衣	83.6

如表2所示,在不同视角下本文提出的特征提取方法在不同多视角的条件下识别率均高于85%。

表2 多视角条件下的识别率

视角/ $^\circ$	18/162	36/144	54/126	72/108	90
全部数据集	94.7	95.5	95.7	96.6	98.5
10张数据集	85.1	85.6	87.6	88.5	88.3

4 结 论

步态识别作为一种远距离的身份识别方式,由于自身的识别特点,可以应用于大多数场合。经实验证明,与传统方法相比,本文基于对抗学习网络框架提出的改进型姿态估计算法提取步态特征对环境变化具有较强的适应能力并且使模型具有更好的泛化能力。同时本文提出的应用于步态识别的特征提取方法突出了对局部细节特征信息的聚焦,能够使模型能够适用于现实中更为复杂的环境之中。

与大多数研究工作者一样,本文采用了图像的方式研

究行人步态,但是随着传感器和监控设备的发展,可以关注更新奇的监控方式,例如红外摄像头,激光雷达等,不同采集方式可以带来更多维度的特征信息。步态识别作为一种身份识别手段,接下来的工作可以研究具体的应用场景例如步态反映的医疗信息,或者在医疗康复领域的研究等。

参考文献

- [1] 何逸炜,张军平.步态识别的深度学习:综述[J].模式识别与人工智能,2018,31(5):442-452.
- [2] DAS C S, TIAHJADI T. Robust view-invariant multiscale gait recognition [J]. Pattern Recognition: The Journal of the Pattern Recognition Society, 2015, 48(3):798-811.
- [3] 邓秀春,王仁芳.基于三维步态特征的步态识别方法[J].计算机仿真,2011,28(4):279-282.
- [4] WU Z, HUANG Y, WANG L, et al. A comprehensive study on cross-view gait based human identification with deep CNNs[Z]. IEEE Trans. Pattern Anal. Mach. Intell, 39(2): 209-226 (2017).
- [5] EL-ALFY H, Xu C, MAKIHARA Y, et al. A geometric view transformation model using free-form deformation for cross-view gait recognition[C]. 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), IEEE, 2017: 929-934.
- [6] CHEN X, KANG Y, CHEN Z. Multi-nonlinear multi-view locality-preserving projection with similarity learning for random cross-view gait recognition[J]. Multimedia Systems, 2020, 26(6): 727-744.
- [7] 胡叶,张爱军.基于K-means的室内行人导航步态识别方法[J].国外电子测量技术,2021,40(3):3236,DOI: 10.19652/j.cnki.femt.2002422.
- [8] CHA Y, KIM H, KIM D. Flexible piezoelectric sensor-based gait recognition [J]. Sensors, 2018, 18(2): 468.
- [9] CHAO H, HE Y, ZHANG J, et al. Gaitset: Regarding gait as a set for cross-view gait recognition [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 8126-8133.
- [10] 贾晓辉,王涛,刘今越,等.基于人体模型映射的步态识别及意图感知方法[J].仪器仪表学报,2020,41(12): 236-244.
- [11] 崔雪红,刘云,常伟,等.基于HOG特征的步态能量图身份识别算法[J].电子测量技术,2017,40(7):100-104.
- [12] 卢来,邓文,吴卫祖.基于改进深度卷积神经网络的步态识别算法[J].电子测量与仪器学报,2019,33(2): 88-93.
- [13] GOWRI H S M, OKADE M. Skeleton based frontal gait recognition utilizing Fourier descriptors [C]. TENCON 2019-2019 IEEE Region 10 Conference (TENCON). IEEE, 2019: 452-456.
- [14] HE Y, ZHANG J, SHAN H, et al. Multi-task GANs for view-specific feature learning in gait recognition[J]. IEEE Transactions on Information Forensics and Security, 2018, 14(1): 102-113.
- [15] ZHANG Z, TRAN L, YIN X, et al. Gait recognition via disentangled representation learning [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 4710-4719.
- [16] LERY V, WEINZAEPFEL P, BREGIER R, et al. SMPLY benchmarking 3d human pose estimation in the wild[C]. 2020 International Conference on 3D Vision (3DV). IEEE, 2020: 301-310.
- [17] LI S, LIU W, MA H, et al. Beyond view transformation: Cycle-consistent global and partial perception gan for view-invariant gait recognition[C]. 2018 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2018: 1-6.
- [18] 章惠,张娜娜,黄俊.优化 LeNet-5 网络的多角度头部姿态估计方法 [J]. 计算机应用, 2021, 41(6): 1667-1672.
- [19] 谭冠政,叶华,陈敏杰.基于频率筛分的无监督人体姿态特征提取与识别研究[J].测控技术,2017,36(9): 7-10,17.

作者简介

李言,硕士研究生,主要研究方向为图像处理。

E-mail:784362548@qq.com

曾维,副教授,主要研究方向为机器学习、图像处理。

E-mail:465630782@qq.com

蒋毅,研究生,主要研究方向为图像处理、嵌入式。

E-mail:529091687@qq.com

王玥妲一,本科生,主要研究方向为图像处理。

E-mail:2490803600@qq.com

罗伟洋,研究生,主要研究方向为图像处理、嵌入式。

E-mail:297045474@qq.com

于真,副教授,主要从事信号与信息处理、数据库技术等方面的研究。

E-mail:605337664@qq.com