基于RGB图像的三维骨骼轨迹的步态识别方法

晏瑶琴,谢 影

上海理工大学计算机与信息工程学院,上海

收稿日期: 2025年1月27日; 录用日期: 2025年2月20日; 发布日期: 2025年2月28日

摘要

步态识别作为一种无接触的人物身份识别方法,近年来受到广泛关注。传统的基于骨骼点的步态识别方 法主要依赖二维图像特征,易受环境因素影响。针对这一问题,本文提出了一种结合RGB图像序列与深 度信息的步态识别方法,以提高识别准确性和鲁棒性。首先,本文使用OpenPose算法从视频帧中提取人 体关键骨骼点,并通过熵度量筛选出对步态识别贡献较大的关节。与传统方法不同,本文将每个骨骼点 的二维坐标转换为三维空间坐标,增强步态特征的表达能力。随后,通过对三维轨迹建模,构建包含时 间序列信息的骨骼点轨迹描述符,并通过计算骨骼点间的相对距离和位移变化,捕捉运动特征。此外, 结合时空信息与深度数据,进一步丰富了骨骼点轨迹描述,构建了具有更高辨识能力的步态特征向量。 为提升识别准确性,本文采用LCSS相似度度量方法,并结合KNN分类器对步态序列进行分类,实现人物 身份识别。

关键词

步态识别,骨架特征,OpenPose,特征融合,KNN分类器

Gait Recognition Method Based on 3D Skeletal Trajectories from RGB Images

Yaoqin Yan, Ying Xie

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Jan. 27th, 2025; accepted: Feb. 20th, 2025; published: Feb. 28th, 2025

Abstract

Gait recognition, as a non-contact method for human identity recognition, has attracted widespread attention in recent years. Traditional skeleton-based gait recognition methods mainly rely on 2D image features, which are easily affected by environmental factors. To address this issue, this paper

proposes a gait recognition method that combines RGB image sequences with depth information to improve recognition accuracy and robustness. First, the OpenPose algorithm is used to extract human key skeleton points from video frames, and an entropy-based measure is applied to select the joints that contribute more significantly to gait recognition. Unlike traditional methods, this paper converts the 2D coordinates of each skeleton point into 3D spatial coordinates to enhance the expressive power of gait features. Subsequently, by modeling the 3D trajectories, a skeleton point trajectory descriptor containing temporal sequence information is constructed, capturing motion characteristics by calculating the relative distance and displacement changes between skeleton points. Furthermore, by incorporating spatiotemporal information and depth data, the skeleton point trajectory description is enriched, resulting in a gait feature vector with higher discriminative ability. To improve recognition accuracy, the LCSS similarity measure is used, and the KNN classifier is applied to classify the gait sequences, achieving human identity recognition.

Keywords

Gait Recognition, Skeleton Features, OpenPose, Feature Fusion, KNN Classifier

Copyright © 2025 by author(s) and Hans Publishers Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/ () ()

Open Access

1. 引言

步态识别作为基于行为分析的生物特征识别技术,已广泛应用于安全监控、身份验证和健康监测等 领域。传统方法多依赖于全局几何特征,如形状轮廓或骨骼结构,来提取步态特征。常见的步态能量图 像 GEI 方法[1],通过分析整体人体轮廓或骨骼几何形态实现个体识别。然而,这类方法在复杂背景中表 现较差,容易受背景干扰及人体形态变化的影响,且对遮挡和姿态变化的鲁棒性较弱,因而在实际应用 中存在显著局限。

随着深度学习和骨骼估计算法[2]的发展,基于骨骼点的步态识别方法逐渐成为主流。这些方法通过 提取人体骨骼关键点,并结合机器学习算法(如支持向量机、KNN等)[3]进行分类,有效捕捉运动特征。 但低质量视频中的误提取和遮挡问题依然挑战识别的准确性。

传统研究主要集中于通过全局几何特征描述步态[4],往往忽视了局部动态模式的变化,尤其是在关 节连接和局部运动方面。Liu 提出了结合时序信息的深度神经网络模型,能够更好地捕捉到步态在不同时 间步的细微变化。Wang 提出了基于 3D 骨骼点轨迹的步态识别方法,通过对骨骼点在三维空间中的运动 轨迹建模,显著提升了对复杂场景中步态的辨识能力,事实上,步态识别不仅依赖于整体步态特征,局 部动态信息(如膝盖、脚踝等关节的运动)对步态节奏和步幅的影响也极为重要。因此,精确捕捉局部运动 变化对于提高识别精度至关重要。现有方法多侧重于短期动态特征,而忽视了长期时空依赖关系,无法 全面表征步态的连续性和复杂性。为了克服传统方法的局限性,本研究提出了一种基于三维骨骼点运动 轨迹的步态建模方法——三维步态轨迹模型(3D-GTM)。该模型通过精确捕捉骨骼点的三维运动轨迹,消 除了对外部环境和地面信息的依赖,有效减少环境干扰的影响。同时,结合动态场景信息生成三维坐标, 并构建全局运动轨迹描述符,关注骨架的局部几何结构和关节间相对位置变化,显著增强了复杂步态中 的辨识能力,通过融合三维骨骼运动轨迹、动态深度信息和局部几何模式,提供了一种更精细高效的步 态特征表示方法,显著提升了复杂场景下的识别表现。

本文的主要贡献如下:

(1) 提出了一种 3D-GTM 模型通过精确捕捉骨骼点的三维运动轨迹进行步态识别,将所有步态特征 嵌入骨骼点的运动轨迹中,避免了传统方法对外部环境或地面信息的依赖,从而减少了环境干扰对识别 准确性的影响。

(2) 提出了一种新的步态全局描述符,用于表示骨架的内部结构和关节的动态运动模式。该方法聚焦 于骨架的局部几何结构,捕捉关节间相对位置的变化及其动态演化。精确表达这些局部运动信息,增强 了步态识别模型的辨识能力。

(3) 该模型通过使用 LCSS 算法进行步态相似度度量,本文解决了时间序列不对齐和噪声干扰问题, 并结合轨迹裁剪技术统一了不同长度视频序列的特征维度。这种方法有效提高了 KNN 分类器对不同步 态序列的处理能力,并保证了输入数据的一致性和稳定性。

2. 模型

2.1. 3D-GTM 模型的架构

图 1 展示了步态识别方法(3D-GTM)图像左侧展示了输入的视频帧序列,作为步态识别系统的原始数据来源。首先,通过 OpenPose 算法[5]从这些帧中提取人体的二维骨骼关键点,生成平面骨骼图(如图中第二部分所示)。与此同时,视频帧输入到深度估计网络[6]计算每个像素的深度值,从而生成深度图像序列(如图中下方所示)。接着,提取的二维骨骼关键点与深度图像数据相结合,推算出人体的三维骨骼点坐标,构建三维骨骼点模型(如图中第三部分所示)。基于三维骨骼点的运动轨迹,提取人体的动态特征,生成全局步态轨迹描述符,用于全面捕捉步态在时间和空间上的动态变化。随后,利用 LCSS 算法[7],计算测试步态与训练样本之间的相似度,从而量化不同步态序列的匹配程度。最终,将 LCSS 相似度度量的特征向量输入 KNN 分类器[8],依据与测试步态最相似的 K 个邻居进行分类分析,输出最终的步态识别结果(如图中最右侧的 KNN 分类器示意图)。整个流程清晰地展现了从数据预处理到步态分类的关键技术与逻辑关系。



Figure 1. 3D-GTM architecture 图 1. 3D-GTM 架构

2.2.3D 骨骼点的构建

首先,本文研究的方法的输入数据可以是专门针对某个行人的行走视频。该视频通常是通过普通摄

像头录制的,可以来自于监控摄像头或固定摄像头,并不要求使用特殊的设备。视频的图像内容为原始 视频帧序列,这些图像帧主要包括一组 RGB 图像序列,这里设步态序列帧为 $S_R = \{s_1, s_2, ..., s_T\}$,其中 T 代表视频序列的图像帧数,即每一帧图像都是标准的彩色视频帧,如图 2 所示。



Figure 2. Illustration of skeleton point extraction using OpenPose algorithm 图 2. OpenPose 算法提取骨骼点的示意图

然后把得到的 RGB 图像序列将 RGB 图像序列输入到 OpenPose 算法中,进行人体骨骼的检测, OpenPose 是一种人体姿态估计算法,通过深度学习模型提取图像中的骨骼关节点,并将这些关节点拼接 成完整的骨骼结构如上图所示。在本研究中,OpenPose 提取了人体的 18 个关键点,这些关键点的二维 坐标(*x*_n,*y*_n)用于表示人体在图像中的空间位置,具体包括鼻子,脖子,左眼,右眼,左耳,右耳,左肩, 右肩,左肘,右肘,左手腕,右手腕,左胯部,右胯部,左膝盖,右膝盖,左脚踝和右脚踝。通过这些关 节点,可以准确地描绘出人体的姿态和运动状态。在步态识别研究中,考虑到与手、脚和头部骨骼结构 的运动相关的主要关节位置,因为这些关节在识别任务中的贡献较大,本章只计算那些对步态识别有高 贡献的关节的坐标。

2.3. 关键骨骼点的筛选

在步态识别中使用熵来量化关节的重要性是一种有效的方法。熵度量[9]通常用于判断关节在运动过 程中的变化幅度和信息量,熵值较低的关节通常表示其在步态过程变化较小的主要特征,因此可以通过 设定阈值来过滤掉这些低熵的关节。对于一个关节J,其熵可以通过以下方式计算。

$$H_{j} = -\sum_{i=1}^{T} P\left(x_{j}^{i}\right) \log P\left(x_{j}^{i}\right)$$
(1)

其中, $P(x_i^i)$ 关节 *j* 在第 *i* 帧的坐标 x_i^i 的概率分布, *T* 是视频序列的帧数。

为了滤除那些对步态识别贡献较小的关节,设定一个阈值 H₀,当关节的熵 H_j小于 H₀时,该关节被 认为是低熵关节,并从后续的计算中剔除。经过计算,本研究选择了 18 个关节中的 13 个关键关节进行 分析,具体包括:右脚、左脚、右膝、左膝、右腕、左腕、右肘、左肘、右手、左手、左胯、右胯和头部。 这些关节的位置信息能有效反映步态特征,增强步态识别的准确性。

在筛选出有效的关节之后,下一步是对步态序列进行处理。具体来说,我们将步态序列帧,在获取到 二维位置坐标,然后把步态序列帧为 $S_R = \{s_1, s_2, \dots, s_T\}$ 输入到深度估计模型中。具体地,每一帧步态序列 都会经过深度估计模型 S_D ,得到相应的深度图像得到相应的深度图像序列 $S_D = \{s_1^D, s_2^D, \dots, s_T^D\}$ 。

最终,结合 RGB 帧和深度帧信息,可以为每个骨骼点生成一系列的三维坐标,对于每一帧图像 S_R 中的骨骼点 (x_n, y_n) 通过与对应的深度帧 S_D 中相同位置的深度值 z_n 进行匹配,将该深度值作为骨骼点的 深度信息 (x_n, y_n, z_n) 。

深度图像中每个点的深度值表示该点与摄像头之间的空间距离。通过深度相机,系统能够获取每个 像素点相对于摄像头的实际距离,这个距离通常是从摄像头到物体表面的直线距离。深度值与物体距离 摄像头的远近呈正相关:距离摄像头越远的点,其深度值越大。

2.4. 关键骨骼点的筛选

通过对每一帧的骨骼点进行上述操作,可以得到每个骨骼点的三维坐标。接下来,为了描述这些骨骼点的运动轨迹,将每个骨骼点的位置信息按照时间顺序连接。具体来说,假设对于第 n 个骨骼点,它在每一帧中的三维坐标为(x_n,y_n,z_n),那么可以得骨骼点在整个视频序列中的三维轨迹:

$$T_n = \{ (x_1, y_1, z_1), (x_2, y_2, z_2) \cdots (x_n, y_n, z_n) \}$$
(1)

为了增强步态描述符的表达能力,可以通过考虑相邻骨骼点之间的相对位置差异来进一步细化运动 轨迹的描述,为了增强描述符的丰富性,考虑相邻骨骼点间的相对位置差异,例如,膝盖与脚踝之间的 距离变化能够反映步态的具体特征,如步幅、步态节奏等。

考虑任意两个骨骼点 *m* 和 *n* 在视频序列的第 *t* 帧中的三维坐标分别为: $T_n(t) = (x_n(t), y_n(t), z_n(t)),$ $T_m(t) = (x_m(t), y_m(t), z_m(t)),$ 其中 $n, m \in \{1, 2, \dots, N\}$ 表示骨骼点的索引, T 是视频是总帧数, 骨骼点 *n*和 m 之间在时间 t 帧的相对距离可以通过欧几里得距离公式计算:

$$d_{nm}(t) = T_n(t) - T_m(t) = \sqrt{\left(x_n(t) - x_m(t)\right)^2 + \left(y_n(t) - y_m(t)\right)^2 + \left(z_n(t) - z_m(t)\right)^2}$$
(2)

其中.表示欧几里得距离, *d_{nm}(t)*表示第 *n* 个和第 *m* 个骨骼点在第 *t* 帧中的相对距离。这一变化度量可以 捕捉到骨骼点在时间维度上的运动信息,反映了骨骼点之间的位移及其动态变化。为了增强描述符的表 达能力,本文将这些相对位移变化信息整合到全局运动轨迹描述符中。具体来说,可以通过以下公式来 表示由骨骼点间相对位置差异构成的全局运动轨迹描述符:

$$D_{nm} = \sum_{t=1}^{T} T_n(t) - T_m(t) + \sum_{t=1}^{T} d_{nm}(t) - d_{nm}(t-1)$$
(3)

这里的 D_{nm} 就是整个视频序列中所有骨骼点的全局运动轨迹描述符。这个描述符综合了每个骨骼点 在空间中的运动变化,并提供了关于人体动作的时空特征信息。在时间维度上将每个骨骼点按照顺序连 接起来,就可以获得由 13 条轨迹组成的一组运动轨迹。这些轨迹分别对应 13 个关节点(例如:右脚、左 脚、右膝、左膝等),每一条轨迹表示该骨骼点在视频序列中的空间运动轨迹,将所有 13 个骨骼点的运动 轨迹组合起来,就形成了一个包含 13 条轨迹的运动轨迹集合,这个运动轨迹集合是步态识别的"全局运 动轨迹描述符",如图 3 所示。



Figure 3. The motion trajectory of human skeletal points 图 3. 人体骨骼点运动轨迹

2.5. 步态识别相似性度量和分类

对于一个包含 t 帧的视频, 通过提取骨骼点的运动轨迹, 可以从视频中生成一系列轨迹描述符

 $(D_{nm}^{1}, D_{nm}^{2}, \dots, D_{nm}^{t-L})$,每个描述符代表个体在特定时间段内的运动特征,从而有效地捕捉运动过程中的时序动态变化。

本文使用非度量 LCSS 算法计算步态识别的相似度,通过动态规划求解最长公共子序列(LCSS),得 到两个时间序列之间的相似度度量,LCSS 方法的核心在于计算两个时间序列之间的最长公共子序列,忽 略小的时间轴偏差,专注于全局相似性。这使得 LCSS 在步态识别任务中具有更强的适应性,尤其在处 理人体运动过程中的局部变化时表现优异。然而,LCSS 方法对时间序列的长度存在一定要求,因为不同 长度的步态数据会导致特征维度的变化,进而影响分类器的准确性。为此,本章通过设定轨迹跨度 L,将 不同长度的步态描述符统一到相同的维度,确保输入数据的一致性。

因为每个输入视频的大小不一致时,即图像帧数不同,会导致从视频中提取出的运动轨迹描述符的 维度也有所不同。为了确保这些不同维度的描述符能够统一并输入到分类器中进行处理,为了适应分类 模型的输入,需要裁剪这些描述符,确保所有的描述符都具有相同的长度。这个裁剪过程涉及设定一个 固定的轨迹跨度 L,将每个视频帧数据裁剪为长度为 L 的轨迹段。对于 t < L 的视频序列,可以通过重复 或者零填充来保证每个序列都有固定长度,对于 L 的取值,在下文中也做了具体的调参实验来确定。

使用 LCSS 算法进行相似度评估之后,采用 k 最近邻平均相似度进行分析,

LCSS_similarity
$$(D_n, D_m) = \frac{LCSS(D_n, D_m)}{\min(n, m)}$$
 (4)

其中 LCSS 相似度计算得出的结果是一个介于 0 和 1 之间的值,表示两个步态序列之间的相似度, LCSS (D_n, D_m) 是这两个序列的最长公共子序列的长度, min(n, m)是两个序列中较短序列的长度。

然后将计算得到的 k 个步态序列 D_1, D_2, \dots, D_k 的 LCSS 相似度度量的特征向量 f_i 输入到 KNN 中,其 中 $f_i = \lceil LCSS(D_i, D_1), LCSS(D_i, D_2), \dots, LCSS(D_i, D_k) \rceil$, D_i 为待识别的步态序列。

然后将测试序列的 LCSS 特征向量 *f_i* 输入到 KNN 分类器中, KNN 算法根据训练数据中与测试样本 最相似的 *K* 个邻居(通过 LCSS 相似度度量)来进行分类。较小的 *K* 值(例如 *K* = 1)容易受到噪声的影响, 导致过拟合, 而较大的 *K* 值则能够平滑分类决策, 增加鲁棒性, 但可能会降低分类精度。一般来说, *K* 值 范围可在 1 到 5 之间进行选择,并通过交叉验证来确定最佳 *K* 值,在下文实验中会有描述。

3. 实验

3.1. 数据集

本章所提出的 3D-GTM 模型(如图 4)在 Gait3D 步态数据集[10]上进行实验,是一个用于步态识别的 大型数据集,特别关注三维空间中的步态分析。它提供了五种类型的数据,2D 姿态,3D 姿态,3D 网格 和光流,该数据集包含了 4,000 位受试者和超过 25,000 个序列,其中 3,000 位受试者的序列用作训练集, 1,000 位受试者的序列 用作测试集。与其他步态数据集相比,Gait3D 特别强调三维空间中的步态,提供 的俯视步态序列能够更好地模拟实际监控摄像头拍摄的情况,具有更高的实用性。

3.2. 参数调整

在步态识别实验中,为了确保不同方法在各个维度上的公平对比,本章参考了大量相关文献和前人 采用的评价方式,选取了业内普遍认可的四种评价指标,用于评估本章模型的性能,并与其他方法的优 劣进行对比分析。这四种评价指标分别是:

(1) 准确率:准确率是指分类正确的样本占总样本的比例,表示模型整体的分类效果,指正确识别的 样本数量与总样本数量之比,常用来衡量系统整体识别效果的好坏。具体的计算公式如下式(6)所示。



Figure 4. Sample data illustration of the Gait3D dataset 图 4. Gait3D 数据集部分数据示意图

准确率 =
$$\frac{TP + TN}{TP + TN + FP + FN}$$
 (5)

(2) 召回率: 召回率表示实际正类样本中,被模型正确识别为正类的比例。衡量步态识别系统在所有可能的步态中,能识别出多少个正确的目标。如果召回率高,说明步态识别系统能够捕捉到更多的目标步态,但可能会有较多的误识别。具体的计算公式如下式(7)所示。

灵敏度 =
$$\frac{\text{TP}}{\text{TP} + \text{FN}}$$
 (6)

(3) 特异性: 是分类任务中用于衡量模型对负类样本(即非目标类样本)正确识别的能力的一个评价指标。它主要用于评估模型对负类样本的排除能力。具体的计算公式如下式(8)所示。

特异性 =
$$\frac{TN}{TN + FP}$$
 (7)

(4) 精度:精度衡量模型预测为正类的样本中,实际为正类的比例。它反映了步态识别预测为正类时的正确性。具体的计算公式如下式(9)所示。

精确度 =
$$\frac{TP}{TP + FP}$$
 (8)

上述的四个指标的数值越高,证明模型的性能越好。在公式(6)到公式(9)中 TP 代表"真正例",是 指模型正确标记为正例的样本数量; FP 代表"假正例",指模型将负类样本错误标记为正类样本的数量; TN 代表"真反例",是指模型正确标记为负类的样本数量; FN 代表"假反例",是指模型将正类样本 错误标记为负类样本的数量。

3.3. 参数调整

在 3D-GTM 模型中,参数调整对模型的性能至关重要。通常需要通过实验来选择最佳的超参数,以 确保模型在不同场景下能够充分发挥其性能。根据 3D-GTM 模型的设计,以下是三个关键的参数,它们 需要通过调参实验来确定:

3D-GTM 模型需要由调参实验确定的参数主要有三个: K 值(KNN 分类器中的邻居数量),骨骼点运 动轨迹描述符的跨度 L (这也是其中最重要的一个参数,它的取值直接影响整个模型的性能)、LCSS 相似 度的阈值 ε。

K值	准确率%	召回率%	特异性%	精度%
1	92.1	90.5	93.7	91.3
2	93.4	91.8	94.5	92.9
3	94.2	93.0	95.0	93.6
4	93.8	92.6	94.8	93.2
5	93.0	91.9	94.3	92.5

 Table 1. Experiment on K-value parameter tuning of the 3D-GTM model on the Gait3D dataset

 表 1. Gait3D 数据集 3D-GTM 模型 K 值参数调整实验

表 1 展示了 K 值(KNN 分类器中的邻居数量)在不同取值下模型的性能指标。实验表明, 当 K=3 时, 模型在 Gait3D 数据集上的整体表现最佳,准确率达到 94.2%,同理,表 2 展示了 L 值的性能指标。

☆ 2. GalGD 数据 案目 距点色列和逻辑处约时度 U 对 I I I I I I I I I I I I I I I I I I				
L 值	准确率%	召回率%	特异性%	精度%
5	90.5	89.2	91.7	90.0
10	93.7	91.8	94.2	93.0
15	94.8	93.9	95.5	94.8
20	94.8	93.6	95.0	93.8
25	92.0	93.9	94.1	92.5

 Table 2. Impact of descriptor span L on performance in the Gait3D dataset

 表 2. Gait3D 数据集骨骼点运动轨迹描述符跨度 L 对性能的影响

当 L = 15 时,模型的整体性能最佳,准确率达到 94.8%。

3.4. 实验结果

实验选取了 20 位被试者,每位被试者提供了 5 条步态序列,定义为 seqi1, seqi2, seqi3, seqi4, seqi5, 其中 i € [1,20],在实验设计中,前 3 条步态序列 seqi1, seqi2, seqi3 用作训练样本,后两条 seqi4, seqi5 用作测试样本,对于训练样本和测试样本,均按照之前提取特征的方法对一个完整步态周期内的每帧图 像进行特征提取,最终得到随时间变化的步态特征向量。表中描述了 3D-GTM 与其他方法在 Gait3D 数 据集上准确率的对比,结果如表 3 所示。

Table	a 3. Experimental	comparison results
主っ	京応オレ佐田米	648

夜 3. 头拉刈比细木奴加	古

方法	准确率%	灵敏度%	特异性%	精确度%
GaitSet [11]	89.38	90.65	91.44	89.62
GaitPart [12]	90.70	90.39	92.01	92.04
GLN [13]	92.66	92.13	93.48	93.92
PoseGait	93.71	93.53	94.11	94.31
ViGait [14]	94.03	93.72	94.90	94.36
3D-GTM	98.68	95.29	95.65	97.32

3D-GTM 凭借 98.68%的准确率,比现有方法提升约 4 个百分点,展现了卓越的步态识别性能。无论 是局部建模的 GaitPart、GLN,还是全局建模的 ViGait,均未达到这一水平,进一步验证了 3D-GTM 的 创新性与实用价值。实验表明,3D-GTM 在多项性能指标上均表现优异,能够高效捕捉关键步态特征, 显著降低误检和漏检,为步态识别领域提供了一种高效鲁棒的解决方案。其在大规模数据集上的出色表现,彰显了技术的推广价值和在实际复杂场景中的应用潜力。

4. 结束语

本文提出了一种创新的步态识别方法——3D-GTM,结合 RGB 图像的深度预测与人体骨骼点的运动 特征,通过深度估计和三维骨骼轨迹建模,显著提升了步态识别的准确性与鲁棒性。3D-GTM 模型通过 结合 OpenPose 算法提取的关键骨骼点与深度估计方法,精确捕捉人体在三维空间中的运动轨迹。相较于 传统步态识别方法,模型能够更好地应对由于时间轴偏移、姿势变化等因素带来的挑战,显著提高了识 别的稳定性和准确性。本研究构建了基于骨骼点相对运动特征的全局步态描述符,并采用 LCSS 相似度 度量进行步态匹配,有效增强了系统对复杂场景与动态变化的适应能力,提高了步态识别的鲁棒性。

在多个标准数据集(如 Gait3D 和 CASIA-B)上进行的实验结果表明,3D-GTM 方法在多视角、不同步态及复杂背景下展现出了优异的识别性能,验证了深度信息与时空特征结合的有效性,本方法在多视角、不同步态和复杂背景下均取得了优异的识别性能。相比于传统方法,3D-GTM 能够有效缓解由于时间轴偏移和姿势变化所带来的挑战,增强了步态识别的稳定性和准确性。

参考文献

- Zou, Z., Chen, K., Shi, Z., Guo, Y. and Ye, J. (2023) Object Detection in 20 Years: A Survey. *Proceedings of the IEEE*, 111, 257-276. <u>https://doi.org/10.1109/jproc.2023.3238524</u>
- Kaur, R. and Singh, S. (2023) A Comprehensive Review of Object Detection with Deep Learning. *Digital Signal Processing*, 132, Article ID: 103812. <u>https://doi.org/10.1016/j.dsp.2022.103812</u>
- [3] Lin, C.J. and Weng, R.C. (2007) Logistic Regression and Support Vector Machines. Proceedings of the International Workshop on Artificial Intelligence and Statistics 2007, San Juan, 21-24 March 2007, 41-47. https://doi.org/10.1109/IWAI.2007.23
- [4] Wang, L., Yang, W.Q. and Zhou, X.Z. (2004) Gait Recognition with Global Geometric Features. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004, Washington DC, 27 June-2 July 2004.
- [5] Wang, L., Yang, W.Q., and Zhou, X.Z. (2004) Gait Recognition with Global Geometric Features. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Washington, 27 June-2 July 2004, 123-130. <u>https://doi.org/10.1109/CVPR.2004.131</u>
- [6] Chen, X., Xu, X. and Zheng, L. (2020) Deep Learning for Monocular Depth Estimation: A Review. *IEEE Access*, **8**, 133974-133986.
- [7] 王海洋,李天佑,张明. 基于 LCSS 算法的时间序列相似性分析研究[J]. 计算机应用与软件, 2021, 38(2): 134-140.
- [8] 陈亮, 刘晓华. 基于 KNN 算法的模式识别研究综述[J]. 计算机科学与技术, 2021, 36(9): 180-185.
- [9] Wang, L. and Zhang, Y. (2020) Research on Signal Processing Methods Based on Entropy. Signal Processing, 36, 112-118.
- [10] Xu, S., Zhang, M., Song, W., Mei, H., He, Q. and Liotta, A. (2023) A Systematic Review and Analysis of Deep Learning-Based Underwater Object Detection. *Neurocomputing*, 527, 204-232. <u>https://doi.org/10.1016/j.neucom.2023.01.056</u>
- [11] Choi, Y., Kim, J. and Kim, S.J. (2019) GaitSet: Towards Good Practices for Deep Learning-Based Gait Recognition. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) 2019, Seoul, 27 October-2 November 2019, 1409-1418. <u>https://doi.org/10.1109/ICCV.2019.00155</u>
- [12] Choi, Y., Lee, H.K. and Kim, S.J. (2020) GaitPart: A Part-Based Model for Gait Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2020, Seattle, 13-19 June 2020, 1027-1035. https://doi.org/10.1109/CVPR42600.2020.01035
- [13] Kim, S.J., Choi, Y. and Kim, J. (2020) GLN: Gait Learning Network for Robust Gait Recognition. *IEEE Transactions on Image Processing*, 29, 7311-7325.

[14] Zhang, X., Yang, S. and Li, X. (2021) ViGait: Vision-Based Gait Recognition Using Transformers. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2021, Nashville, 20-25 June 2021, 305-313. <u>https://doi.org/10.1109/CVPR46437.2021.00305</u>