

可穿戴计算中基于步频分布的身份识别研究

王彬 付雄* 王俊昌

(南京邮电大学计算机学院 江苏 南京 210023)

摘要 针对目前可穿戴计算领域中运用生物识别技术进行身份识别产生的低效、设备能耗大的问题,提出步频分布的新型步态特征,并利用步频分布特征提出基于相对熵的身份识别算法。利用加速度传感器对步态数据进行采集,建立步频分布的特征模型,利用相对熵判别用户身份。实验结果表明,该算法的识别准确率达到 86% 以上,有效降低了设备能耗,提高了实际应用价值。

关键词 可穿戴计算 身份识别 相对熵 加速度传感器

中图分类号 TP391

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2020.10.030

IDENTIFICATION BASED ON STEP FREQUENCY DISTRIBUTION IN WEARABLE COMPUTING

Wang Bin Fu Xiong* Wang Junchang

(School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, Jiangsu, China)

Abstract In view of the inefficiency and high energy consumption of biometric identification technology in the field of wearable computing, a new gait feature of step frequency distribution is proposed. Based on the characteristics of the step frequency distribution, we propose an identity recognition algorithm based on relative entropy. The gait data was collected by the acceleration sensor, and then the feature model of the step frequency distribution was established. The relative entropy was used to determine the user identity. The experimental results prove that the recognition accuracy of our algorithm is more than 86%, which effectively reduces the energy consumption of the equipment and improves the practical application value.

Keywords Wearable computing Identification Relative entropy Acceleration sensor

0 引言

随着可穿戴技术的飞速发展,可穿戴设备已经在人们的生活中越来越普遍,但是在设备所有者隐私保护方面尚未形成有效且标准的做法^[1]。为了使各种可穿戴设备中的重要信息不外漏或不被盗取,需要可穿戴产品能够实现身份识别的功能。目前主流的口令、密码等认证方式虽然能有效地认证身份,但存在着遗忘和丢失的问题,同时还需要用户主动配合,这通常会带来不必要的麻烦和损失^[2]。

近年来,利用生物特征进行身份识别已经成为计

算机领域和生物医药科学领域中的首选安全验证方式^[3]。步态特征是生物行为特征中的一种典型特征,主要是指人在正常步行状态下所表现出的行为特征,它不易被盗取而且广泛存在^[4]。步态特征差异通常表现在步频、步长、脚落地时的轻重程度等几个方面。利用人步行状态下独有的特点能够有效地区分个体^[5]。

目前基于步态特征的身份识别技术主要有以下三个方向:机器视觉、地面压力传感器和加速度传感器。机器视觉的方法首先利用图片拍摄或视频录制等手段采集用户行走时的图像信息建立个人信息库,再通过图像识别技术进行比对从而辨别不同的用户^[6],但其存在着视觉范围、障碍物遮挡、拍摄角度、光线不足等

问题,继续发展和推广应用的局限性较大。基于地面压力传感器的方法在地面上铺置压力传感器,然后采集用户步行通过时足底压力的分布情况^[7],但该方法需要在特定设备和场景下才能进行有效的身份验证,实际中应用较少。加速度传感器具有不受视觉影响和场地限制、数据采集方便快捷的优势,因此基于步态特征的身份识别更多是依赖加速度传感器实现。

近些年来,国内外的专家运用加速度传感器采集步态特征用于步态识别出现了很多方向。Derawi 等^[8]采用时间插值固定采样点的时间间隔、平均周期长度,以步态样本的幅值序列和时间序列为特征,利用动态规整算法(Dynamic Time Warping, DTW)计算不同长度序列的距离,识别率有所提高,但是需要在数据处理阶段进行大量工作,包括数据归一化、数据样本维数调整,且在识别阶段需要花费大量时间用于规整路径,这对于计算能力有限的移动终端来说是不可接受的。

2019年,孔菁等^[9]对手机自定义坐标轴进行空间转换,选用极大值、均方根、梅尔频率倒谱系数等7种特征,使用支持向量机算法来对样本进行训练分类,达到身份识别的目的。运用机器学习算法需要较高的配置,会带来高能耗的问题,而且该算法高能耗换来的准确度提升也不明显。2019年,Sun 等^[10]提出一种速度自适应步态周期分割方法和个性化阈值生成方法,采用基于快速傅里叶变换(FFT)的步态周期估计方法来提高步态分割的准确性,使用皮尔森相关系数来衡量步态周期模板和用户步态周期的相似接近程度。与具有固定速度和阈值的可用方法相比,在各种步行速度下的步态和用户识别率得到一定改善。

然而上述研究都是基于恒定步行速度的假设,而且都是在划分出步态周期后,再应用基于距离或度量函数的方法如 DTW、PCC (Pearson Correlation Coefficient) 进行识别,这就需要明确的步进周期检测,会受周期检测失败和周期性相位错位的影响^[11]。在实际应用场景中,会出现步行异常导致步态周期的不稳定,这时单个步态周期在时域上不同,通过插值等手段度量单个步态周期频域上的序列距离作为认证标准具有片面性。

鉴于此,本文不针对单个步态的差距进行深入讨论,着重于用户行走习惯上的差异,提出步频分布的新型步态特征,进行持续认证下的步态周期分布差异研究,从而达到身份识别的目的。文献[12]的研究表明,行人的步频、步速同为 Gauss(正态分布),平均步频为 1.96 step/s,而且步频分布在不同人群中的差异性比较大,这也揭示了步频分布用于身份识别的可行性。

本文提出一种基于步频分布的身份识别算法,主要有两部分组成:步频分布模型的刻画、身份识别算法。首先在数据处理阶段对加速度传感器的数据进行采集;然后在特征提取与建模阶段通过动态阈值切割步态,统计得到步频分布建立步态特征模型;最后基于相对熵得到识别结果。实验结果表明,基于步频分布的身份识别算法的识别准确率达到 86% 以上,可在大样本认证情况下有效地识别用户身份,对比当前的 DTW 和 PCC 算法,设备能耗最多不超过 DTW 和 PCC 算法的 50%,提高了步态识别的准确度,降低了可穿戴设备终端的能耗。

1 特征提取和建模

正常步态情况下,从一侧脚跟着地开始到该侧脚跟再次着地为止的时间被称为一个步态周期。一个完整的步态周期分为首次触地、支撑和摆动三个阶段,每个阶段对应不同步态时期。步频是人行走 1 s 的步数。通过计算一个步态周期的时间 SingleStepTime 来作为一步的时间,则步频 StepFrequency 需要用完整步态周期的持续时间的倒数来表示。为简化,本文直接使用一个完整步态周期的时间来作为刻画步频的一种方式,下文谈论步频分布都指的是单个步态时间的分布。StepFrequency 的计算如下:

$$\text{StepFrequency} = \frac{1}{\text{SingleStepTime}} \quad (1)$$

本文提出一种基于步态切割的步频分布生成算法,基本思想为:分析出人在行进过程中,加速度传感器的信号呈现周期性的特征,以波峰为界限,切割传感器信号获得单个步态,并统计单个步态周期的时间记录入步频分布。相关变量的定义如表 1 所示。

表 1 变量和函数定义

变量	定义
P	加速度信号集合
P_i	加速度集合中第 i 个信号量
S	步频分布集合
S_i	步频分布中第 i 个区间值
T_p	上一次波峰时间
C_p	当前波峰时间
T	当前系统时间
p	波峰值
v	波谷值
α	波峰阈值

续表 1

变量	定义
β	两个波峰时间差阈值
μ	波峰波谷差值阈值
d	步态时间
$flag$	上升趋势标志位
$count$	上升次数

首先对步频分布区间进行处理,人行走一步的时间约为 $0.4 \sim 0.6 \text{ s}^{[13]}$,在区间 $[0.4, 0.6]$ 之间作 m 等分,则每个单独步态都会分布在其中的小区间上,初始每个子区间为 0。

接着对持续产生的加速度信号需要进行波峰的检测,波峰需要满足的条件:

- (1) 目前的信号 P_i 为下降趋势;
- (2) 之前的信号 P_{i-1} 为上升趋势;
- (3) 持续上升的次数等于 2;
- (4) 信号值大于上阈值 α 。

在检测到波峰之后,为了避免噪声导致的“伪波峰”的影响,需要过滤掉波峰相距太近的情况,要求两个相邻波峰时间差大于两个波峰差的阈值 β 。在行进过程中,不断地梯度化更新波峰波谷的差值阈值 μ ,确保可以过滤掉噪声的同时,不会遗漏掉有效数据。最后计算两个相邻波峰的时间差值 d ,遍历步频分布区间 S 找到对应子区间 S_i ,频数加 1。具体算法如算法 1 所示。

算法 1 基于步态切割的步频分布生成算法

输入: 加速度信号集合 P 。

输出: 步频分布序列 $S = \{S_1, S_2, \dots, S_m\}$ 。

1. 初始化 S 每个区间
2. for each P_i in P do
3. if $P_i > P_{i-1}$ then
4. $flag = true$
5. $count ++$
6. else
7. $flag = false$
8. $count = 0$
9. if $!flag$ and $count = 2$ and $P_i > \alpha$ then
10. $p = P_{i-1}$
11. $T_p = C_p$
12. $T =$ 当前系统时间
13. if $T - T_p > \beta$ and $p - v > = \mu$ then
14. $C_p = T$
15. $d = C_p - T_p$
16. 找到 d 所在的区间 S_i
17. $S_i ++$

18. end if
19. if $(T - T_p) > \beta$ then
20. 梯度化阈值 μ
21. end if
22. else
23. $v = P_{i-1}$
24. continue
25. end for
26. return S

算法在初始化步频分布集合 S 时,划分为 m 个小区间,时间复杂度为 $O(m)$;加速度信号集合 P 有 n 个数据,遍历 P 的时间复杂度为 $O(n)$,判断波峰需要借助常数级的空间,空间复杂度为 $O(1)$;寻找单个步态落到的步频分布区间,时间复杂度为 $O(m)$,算法整体时间复杂度 $O(m \times n)$ 。

2 基于相对熵的身份识别算法

相对熵,又称为信息散度或 KL 散度,源于信息论,是衡量概率分布相似度的一种方式,表示一条信息的信息量大小和它的不确定性有密切关系。信息领域中,相对熵常用来量化两个随机分布之间的距离,当这两个随机分布相同的时候,它们的相对熵为 0,当两个随机分布的差异越大时,它们之间的相对熵也会越大。

根据第 1 节基于步态切割的步频分布生成算法,可以建立不同人物的步态特征模型,判断两份步态特征模型是否为同一个人是身份识别的最终目标。本文在识别阶段,首先对步频分布处理,获取步频的概率分布,然后计算两份步频概率分布的相对熵来比较两份步态数据的相似度,从而达到身份识别的目的。

假设对于一个未认证的认证者 B ,步频分布为 $B(x) = \{B_1, B_2, \dots, B_n\}$,其声称与目标用户 A 为同一个人。那么需要将认证者 B 的步频特征与用户 A 的步频特征进行比较,假设用户 A 的一份步频分布为 $A(x) = \{A_1, A_2, \dots, A_n\}$,可以通过式(2)、式(3)计算 A 、 B 用户的两份样本的步频概率分布 $P_A(x) = \{a_1, a_2, \dots, a_n\}$ 、 $P_B(x) = \{b_1, b_2, \dots, b_n\}$ 。

$$a_i = \frac{A_i}{\sum_{j=1}^n A_j} \quad (2)$$

$$b_i = \frac{B_i}{\sum_{j=1}^n B_j} \quad (3)$$

可以确定样本 $B(x)$ 中的步频分布信息的不确定度为 $\sum_{j=1}^n a_j \log \frac{1}{b_j}$ 。同理,若 $B(x)$ 和 $A(x)$ 属于同一个人

的两份样本,那么可得样本 $B(x)$ 中的步频分布信息的不确定度为 $\sum_{j=1}^n a_j \log \frac{1}{a_j}$ 。

因此,样本 $A(x)$ 和 $B(x)$ 的差异为:

$$\sum_{j=1}^n a_j \log \frac{1}{b_j} - \sum_{j=1}^n a_j \log \frac{1}{a_j} = \sum_{j=1}^n a_j \log \frac{a_j}{b_j} \quad (4)$$

式(4)计算得到的值,代表了概率分布 $P_A(x)$ 与 $P_B(x)$ 之间的差异的一个度量。然而通过一份样本进行识别误差很大,考虑到之前 A 用户会有多份样本存在,把之前的样本也参与比较分析会提高识别准确率。

对于用户 A 已有的样本 $A_1 = \{A_{11}, A_{12}, \dots, A_{1n}\}$ 和 $A_2 = \{A_{21}, A_{22}, \dots, A_{2n}\}$ 是同一个人的两份样本,可以计算 A_1 和 A_2 的相对熵:

$$D_{kl}(A_1 \parallel A_2) = \sum_{j=1}^n a_{1j} \log \frac{a_{1j}}{a_{2j}} \quad (5)$$

式(5)可以应用于合法用户 A 的所有样本 $\{A_1, A_2, \dots, A_n\}$,得到合法用户 A 的平均相对熵:

$$D(A) = \frac{2 \sum_{i=1}^n \sum_{j=i+1}^n D_{kl}(A_i \parallel A_j)}{n(n+1)} \quad (6)$$

对于一个声称用户 A 的用户 B ,应用式(4),对于任意用户 A 的样本 $A_i (i=1, 2, \dots, n)$,如果满足:

$$D_{kl}(B \parallel A_i) \leq k \times D(A) \quad (7)$$

那么认定 B 是 A 的一个样本,认证成功;否则,认定 B 是一个伪冒者。

本文采用相对熵衡量两份概率分布的相似性。在实际应用过程中,很容易出现样本某个子区间的步频为 0,导致计算 KL 散度因为出现子区间的步频概率为 0, $\lim_{p \rightarrow 0} p \log p = 0$ 。然而当 $p \neq 0$ 且 $q = 0$ 时,应用式(5), $D_{kl}(p \parallel q) = \infty$,这样就断定了两份样本不是同一个人,针对这种极端情况。本文采用如下的方式计算相对熵:

假设样本 B 声称是 A 用户,两份步频概率分布 $P = (a:3/5, b:1/5, c:1/5)$, $Q = (a:5/9, b:3/9, d:1/9)$,此时引入常量 τ ,然后可以将 P 和 Q 转换为 $P' = (a:3/5 - \tau/3, b:1/5 - \tau/3, c:1/5 - \tau/3, d:\tau)$, $Q' = (a:5/9 - \tau/3, b:3/9 - \tau/3, c:\tau, d:1/9 - \tau/3)$,并求 $D_{kl}(P' \parallel Q')$ 。

3 实验

本文提出的可穿戴计算中基于步频分布的身份识别算法,用于在可穿戴终端上能够高效率低功耗地识别终端用户。实验选用 Android 系统平台来评估本文的算法。在安装开发好的应用程序之后,以后台服务

形式运行在宿主机上,无须用户手动干预。硬件环境为华为荣耀 V9,操作系统为 Android 7.0。

为了体现本文算法的优越性,我们还与另外两种经典的步态识别算法进行对比测试。一种是基于 DTW 的步态识别算法^[8],该算法使用频域上的序列作为步态特征,使用 DTW 算法规整路径,计算欧式距离来实现身份识别,本文简称这种算法为 BDI(Base on DTW Identification);另一种是基于皮尔森相关系数相关相似性的模板匹配算法(PCC)^[10],本文简称这种算法为 BPI(Base on PCC Identification)。本文提出的算法简称为 BSDI(Base on StepFrequency Distributed Identification)。

3.1 实验数据采集

在实验数据采集阶段,利用 Android Studio 开发工具开发一个小型的应用程序进行数据采集,设置采样频率 60 Hz,并实时记录数据到数据库中。实验平台内置三轴加速度传感器,三轴加速度传感器的坐标系是以手机屏幕为基准的相对坐标系^[14]。以手机的几何中心为原点,X 轴平行于手机机身的短边且以向右侧为正方向,Y 轴平行于机身的长边且以向听筒方向为正方向,Z 轴垂直于手机机身且以屏幕的朝向为正方向^[15]。不同方向的加速度信号值如图 1 所示。

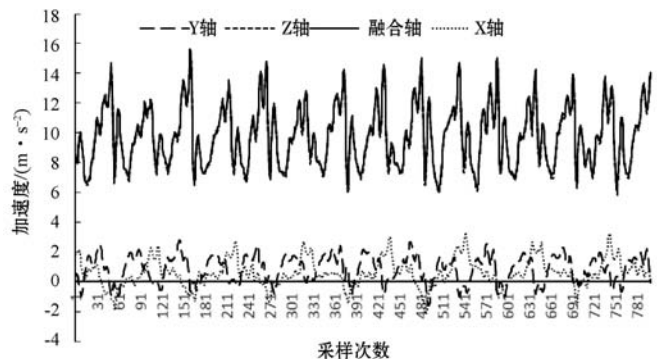


图 1 加速度信号图

在选取数据源时,由于 Y 轴和 X 轴在用户行走时不太敏感,而 Z 轴和融合轴 M(Mix)的加速度呈现较大的周期性特征^[16],本文采用三轴加速度融合之后的融合轴 M 作为实验数据。融合轴 M 的信号值计算为:

$$M = \sqrt{X^2 + Y^2 + Z^2} \quad (8)$$

3.2 实验过程和结果分析

实验采集阶段共采集了不同身高、不同性别、不同年龄段共 25 名志愿者(编号 1-25)两天的步态数据,每名志愿者每天手持手机步行时间大约 20 min(去除起始与结束阶段的不稳定数据),统一在摩擦系数等客观因素均相同的操场跑道进行步行。共采集了 500 ($2 \times 25 \times 10$)份样本。单个样本为步行 200 步在 $[0.4,$

0.6] 的 20 个小区间的步频分布。每个志愿者步行前一天的数据用于训练,计算该志愿者在式(7)中的认证阈值 k 。后一天用于后续实验测试部分。

如图 2 所示,本文选取了志愿者 1-4 展示步频特征的差异性,显而易见,同一个志愿者的不同样本的分布拟合度很高,而不同志愿者的差异很明显。

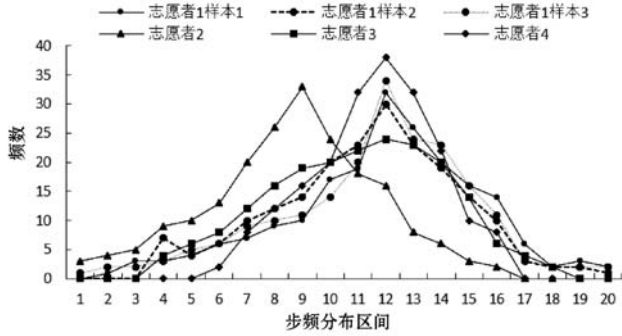


图 2 不同志愿者步频分布图

如图 3-图 4 所示,实验测试过程中,数据表明在计算相对熵衡量两份样本的相似度时,伪冒认证的相对熵大都大于 0.2,而真实认证的相对熵稳定在 0~0.15 之间。

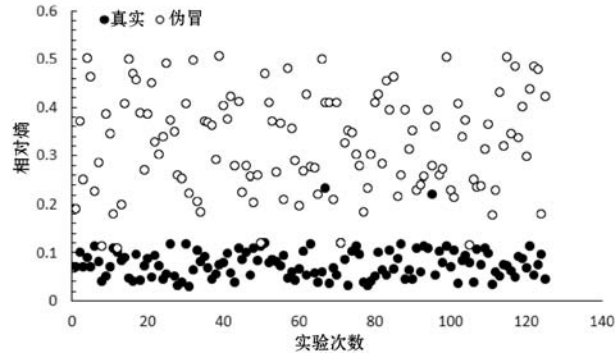


图 3 志愿者 1 的样本认证相对熵对比图

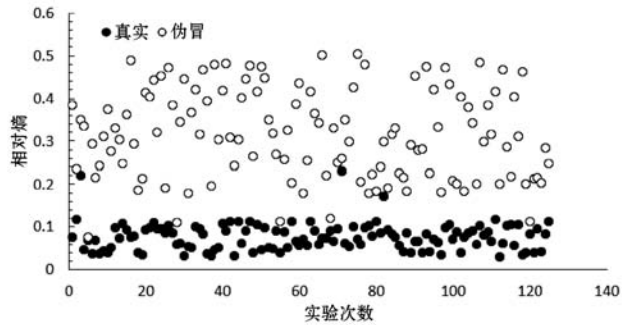


图 4 志愿者 2 的样本认证相对熵对比图

图 3 和图 4 中,真实指的是同一个志愿者的不同样本进行的交叉认证,伪冒是其他志愿者的样本分别来认证志愿者 1 和志愿者 2。

实验训练阶段利用 25 名志愿者的前一天的样本训练,每个志愿者的 10 份样本交叉认证,令 FRR(错误拒绝率)为 0,计算得到 25 名志愿者在式(7)中的认证阈值 k ,结果如表 2 所示。

表 2 实验结果

目标用户	上阈值 k	错误认证数	准确率/%
1	1.368	28	88.8
2	1.455	25	90.0
3	1.652	29	88.4
4	1.586	26	89.6
5	1.436	24	90.4
6	1.388	22	91.2
7	1.644	27	89.2
8	1.438	20	92.0
9	1.630	23	90.8
10	1.685	25	90.0
11	1.623	24	90.4
12	1.535	21	91.6
13	1.297	12	95.2
14	1.552	17	93.2
15	1.414	16	93.6
16	1.495	17	93.2
17	1.358	15	94.0
18	1.418	18	92.8
19	1.533	19	92.4
20	1.669	27	89.2
21	1.410	21	91.6
22	1.602	26	89.6
23	1.394	17	93.2
24	1.530	24	90.4
25	1.354	17	93.2

实验测试阶段,将 25 个志愿者进行 1-25 编号,取 1 号志愿者作为目标用户,选取 1-25 号每个志愿者的 10 个样本进行认证(以此类推到 25 号志愿者),共计进行实验 $(25 \times 25 \times 10) = 6250$ 次,识别结果如表 2 所示。

为验证本文算法的性能和效率,将其与 BDI、BPI 两种步态识别算法在两个方面进行对比实验:不同样本量情况下识别率的对比,不同时间下持续认证给设备带来的能耗对比。实验时 BDI 和 BPI 均采用 60 Hz 的采样率。

以 25 个志愿者作为目标人物,重复实验取平均值,结果如图 5 所示。逐渐增加伪冒认证和真实认证的样本量,本文算法仍然可以保持 86% 以上的识别准

准确率,而其他算法识别准确率在随着样本量增大之后逐渐降低。真实样本数是选取作为目标用户的样本,伪装样本是其他志愿者的样本。

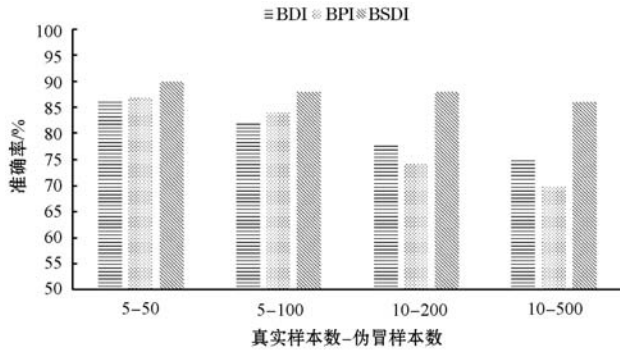


图5 不同样本量下的识别率比较

进行能耗对比实验时,实验平台荣耀 V9 手机所有的前后台应用关闭,仅运行本文的身份识别应用程序,采用耗电量(mAh)作为能耗的体现。如图6所示,本文算法无论是短时间的认证,还是长时间的认证,耗电量都很低,最多仅有其他两种算法的50%。

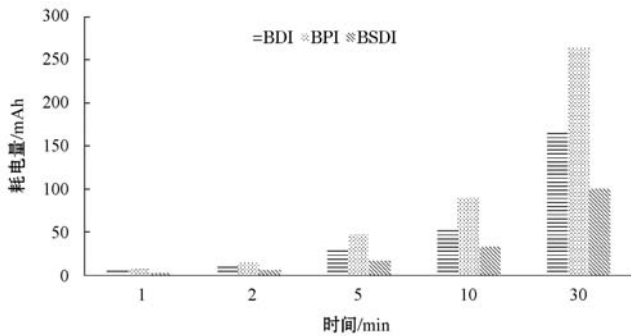


图6 不同时间下持续性认证的能耗比较

由上述两方面的对比实验证明,本文算法在大样本认证情况下,具有较高的识别准确率且能耗低。这表明,本文提出的基于步频分布的身份识别算法能有效地识别身份,并可用于计算能力较弱、内存小的穿戴式设备中。

4 结 语

本文针对现阶段可穿戴计算领域中身份识别算法的低效和能耗大的问题,在充分分析步态构成之后,着重考虑用户的行为习惯,提出一种基于步频分布的身份识别方法,刻画一种新型的步态特征,并通过相对熵进行认证。实验证明,本文算法有着较高的识别准确率和较低的能耗,在各种可穿戴设备上会有比较好的实际应用价值。虽然本文算法准确率比较高,但是在步态切换时不能表现出良好的识别度,未来将继续改进本文算法,使其在实际应用中有更高的实用性。

参 考 文 献

- [1] 刘强,李桐,于洋,等. 面向可穿戴设备的数据安全隐私保护技术综述[J]. 计算机研究与发展,2018,55(1):14-29.
- [2] 郑方,艾斯卡尔,王仁宇. 生物特征识别技术综述[J]. 信息安全研究,2016,2(1):12-26.
- [3] 杨鹏. 步态识别中的常用方法[D]. 长春:吉林大学,2010.
- [4] 王亮,胡卫明. 基于步态的身份识别[J]. 计算机学报,2017,26(3):353-360.
- [5] 叶汉民,黄培亮. 基于下肢动静态特征融合的步态识别[J]. 科学技术与工程,2016,16(8):218-222.
- [6] 刘冠群,罗桂琼,谭平. 基于类 Haar 特征模板匹配的多镜头步态识别算法[J]. 计算机工程,2017,43(12):231-236.
- [7] 佟苏洋,和焕胤,汤澄清. 基于足底压力分析系统对青年人行走步态特征稳定性的研究[J]. 四川警察学院学报,2017,29(6):77-83.
- [8] Derawi M O, Nickel C, Bours P, et al. Unobtrusive user-authentication on mobile phones using biometric gait recognition [C]//2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing,2010.
- [9] 孔菁,郭渊博,刘春辉,等. 基于智能手机运动传感器的步态特征身份识别方法[J]. 计算机应用,2019,39(6):1-9.
- [10] Sun F, Mao C, Fan X, et al. Accelerometer-Based speed-adaptive gait authentication method for wearable iot devices [J]. IEEE Internet Of Things Journal, 2019,6(1):820-830.
- [11] Zhang Y, Pan G, Jia K, et al. Accelerometer-Based gait recognition by sparse representation of signature points with clusters[J]. IEEE Transactions On Cybernetics, 2015, 45(9):1864-1875.
- [12] 陈然,董力耘. 中国大都市行人交通特征的实测和初步分析[J]. 上海大学学报(自然科学版),2005(1):93-97.
- [13] Krzeszowski T, Switonski A, Kwolek B, et al. DTW-Based gait recognition from recovered 3-D joint angles and inter-ankle distance[C]//International Conference on Computer Vision and Graphics,2014.
- [14] Muhammad N B. A study on cell phone security: Authentication techniques [C] //International Conference on Information and Computer Technologies,2018.
- [15] Harbach M, Luca A D, Egelman S. The anatomy of smartphone unlocking: a field study of android lock screens [C] //CHI Conference on Human Factors in Computing Systems, 2016.
- [16] Hoang T, Nguyen T. Adaptive cross-device gait recognition using a mobile accelerometer[J]. Journal of Information Processing Systems,2013,9(2):333-348.