软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2019,30(3):700−717 [doi: 10.13328/j.cnki.jos.005687]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

面向通用模型的高可用性步态周期分析方法*

门慧超, 王波涛

(东北大学 计算机科学与工程学院,辽宁 沈阳 110819)

通讯作者: 王波涛, E-mail: wangbotao@cse.neu.edu.cn

摘 要: 步态数据分析是模式识别、数据挖掘与智能数据分析领域中的一个重要研究方向.对步态数据进行分析并求解步态周期段,是该领域的一个核心课题,其中,以波峰波谷识别、步态模板匹配、利用信号处理基本方法为主.这些方法尽管已取得了一定条件下的应用,但大多需要预知步态数目、步态模板数据等基本信息,其可用性受限.针对这一问题,提出一种结合了波峰波谷检测与阈值空间的高可用性步态周期分析方法,通过自动求解预估值并构建自适应区间,根据通用步态模型对缺乏上述信息的未知步态数据进行切分与分析,能够更便利准确地求解步态周期数据.同时也提出了一种过滤不相关数据的方法,用于过滤测试数据头尾的部分不相关数据.该方法与朴素快速傅里叶方法和本领域3类最新典型方法进行了对比,实验结果显示:在所有样本中,该方法的步态周期段求解准确度最高,使步态数据的分析与处理更贴合实际情况.

关键词: 步态周期数据;步态分析;智能数据分析;模式识别;数据挖掘

中图法分类号: TP18

中文引用格式: 门慧超,王波涛.面向通用模型的高可用性步态周期分析方法.软件学报,2019,30(3):700−717. http://www.jos.org.cn/1000-9825/5687.htm

英文引用格式: Men HC, Wang BT. High feasible method of cycle segment analysis for common gait model. Ruan Jian Xue Bao/Journal of Software, 2019,30(3):700−717 (in Chinese). http://www.jos.org.cn/1000-9825/5687.htm

High Feasible Method of Cycle Segment Analysis for Common Gait Model

MEN Hui-Chao, WANG Bo-Tao

(School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China)

Abstract: Gait analysis is an important research in the fields of pattern recognition, data mining, and intelligent data analysis. Many different methods, such as recognition of crests and troughs, matching of gait templates, signal-processing based approaches, etc., have been proposed and applied, but these methods require presetting parameters such as gait numbers, gait templates, etc. Their feasibilities are limited. In this study, a new high feasible method, which can automatically calculate a presetting cycle value and a self-adaption range, is proposed, based on peak detection and threshold, to analyze gait cycle segment without unnecessary information needed by other methods. Moreover, a method to filter nonrelated data is proposed. Comparing with simple FFT and other three kinds of representative methods, the proposed approach has the best analysis results in evaluations.

Key words: gait cycle segment data; gait analysis; intelligent data analysis; pattern recognition; data mining

步态数据分析是模式识别与数据挖掘领域的一个重要课题,它代表了人类在走路这一最常见活动中的行为,通过一定智能的方法分析步态数据,得到许多有价值信息,进而可以将分析后的此类信息应用到各种领域,例如体育领域中提高运动员的表现,提高训练质量和健康判断[1];在医学领域监控病人复健进展,评估病人及老

* 基金项目: 国家自然科学基金(61672144, 61173030)
Foundation item: National Natural Science Foundation of China (61672144, 61173030)
本文由智能数据管理与分析技术专刊特约编辑樊文飞教授、王国仁教授、王朝坤副教授推荐.
收稿时间: 2018-07-19; 修改时间: 2018-09-20; 采用时间: 2018-11-01

年人的日常生活风险,协助对他们的未来患病的预测等[2].另外,在生物识别和生物工程领域也有广泛的应用[3].

步态数据的本质是一组周期性的时间序列数据,在时间序列数据的数据挖掘与分析领域,周期性、季节性、趋势性等是时间序列数据的基本特征[4],通过这些特征,可以描述时间序列数据的许多信息.步态数据切分的思想即是利用步态数据作为时间序列数据的这一本质,利用时间序列数据的基本特征来对数据进行分析进而切分,进而可以进行步态识别,或更深入扩展到动作识别领域.

步态数据分析的常用数据源主要是加速度计与陀螺仪两种设备,其中,陀螺仪测定的角加速度在较短时间内比较准确,而较长时间会因存在漂移而有误差[5−7];加速度计测定的线加速度在较长时间的测量值是正确的,在较短时间内由于信号噪声的存在而有误差[8−10].二者各具特点,可根据不同研究或工程需求选用,或同时使用二者结合的传感器作为数据源[11],均取得了较好的实验成果.另外还有部分其他不常见的数据源,如脚踏板开关信号[12]、压力计[13,14]等.本文将以线性加速度作为数据源,只使用单一数据源即可以更好地达到对步态数据解析与分析并进行切分的目的.

目前,在步态数据分析领域有大量文献成果.

- 文献[5]采用陀螺仪数据源结合阈值的思想进行分析,但阈值多为作者人工定义常数,同时并未给出未知步态步数的解决方案;
- 文献[6,14]均采用马尔科夫链进行建模,其中:
  - 文献[6]是一篇针对特殊疾病患者的交叉学科文献,其采用陀螺仪采集的角速度数据,结合马尔科夫模型进行分析,与本文课题无论是数据源类型还是整体角度抑或是研究思路均不相同.其思路为:将步态数据分为不同类型的模型,再分别根据多种不同模型建立马尔科夫链,进而对患者情况进行判断.研究目的是对不同类型的步态模型进行对比,进而找到最佳的模型(model);
  - 文献[14]是一篇针对步态识别的文献,采用压力计作为数据源,并建立马尔科夫模型进行分析.应用马尔科夫模型进行模式识别是一种常用的方法,此方法的特点是可应用范围较广、使用灵活,但计算较为复杂,需要建立较为复杂的非线性训练模型,没有固定的解决方法,需选择合适的状态数和特征矢量的维数.为了避免这些缺点,提高识别便捷性,本文采用了基于规则的分析识别方法;
- 而文献[11]则同时采用了两种数据源,结合了线加速度与角加速度对步态数据进行分析与处理.这是一种较为成熟的方法,但需同时具备陀螺仪与加速度计两个数据源条件,在一些情况下,实验条件可能并不具备两个数据源,此时,此方法就有了局限性;与此同时,此方法在对步态段数据进行标记与检验这一关键阶段,依然以一个人工定义的恒定常数阈值作为后续的基础,与文献[5]具有相似的缺陷;
- 文献[13]采用的是压力计,同时提出了一种基于 K-多项式分布来检验步态段之间间隔的方法.这是一种与本文思路区别较大的方法,该方法从数学角度解决了步态段的识别问题,其特点是方法较为科学扎实,但计算较为复杂;
- 文献[12]采用的是一种脚踏板开关信号作为数据来源,这种信号是离散型的,但其设备较为不常见,而采用线加速度计及陀螺仪产生的连续数据必须进行离散化处理,才有一定可能性使用该文献的算法;
- 文献[7]采用的也是陀螺仪数据源,其方法由于设定了特定的恒常数阈值,在该文献数据集内效果较好,但不一定能够较好地适用于其他数据集.

综上所述,在步态数据分析领域有许多研究方法,主要分为以下几类.

(1) 基于峰值谷值侦测的方法.这类是最基础的方法,对于较为规律的步态数据,采用此种方法可以得到较好结果[8,15],许多进行步态分析的工程项目或实际应用都采用这类方法.但其缺点是由于其标准只依赖于波峰波谷,遇到较为复杂的步态情况时很容易因为噪声或其他因素影响造成误判,譬如在实际生活中,经常出现同一段计步数据不同软件记录到不同步数的情况(此时除了传感器方面的原因,也因为软件计步方法不同而有所区别).文献[8]是此类方法的一个代表性方法(在第 2.1.1 节进行了详述,并在后文进行了对比实验);

(2) 基于步态模板数据(template)与峰值阈值相结合的方法[9,16,17].这类方法的优点是:由于已求解或保存了模板数据,切分识别率更高.但由于模板数据可能是固定存在或通过某一类单一步态数据段计算而来,局限性较大,也面临着无法动态识别不同情况下可能出现的起伏幅度较大、变化更为灵活的步态的问题.文献[9]作为一个代表方法在第 2.1.2 节进行了详述,该文献后续的实验也证明此方法效果不佳;同时,此文献的方法要求必须预知步数信息,且必须提前计算出一组或几组步态模板数据,应用较为局限,无法处理未知数据;另外,此文献采用的步态模型与本领域普遍研究的通用模型不同,也造成了该文实验结果不尽如人意;

(3) 基于朴素傅里叶变换的方法[18].这是因为步态数据作为时间序列数据的这一本质,但由于其求解的周期是固定常数,单独使用这一固定常数很难有效地对步态数据进行切分与分析,数据中后期偏移会逐渐增大.本文第 2.1.3 节对这一方法进行了简述,并在后文进行了对比实验;

(4) 前 3 点均为基于规则的(rule-based)识别分析方法,在相关研究中使用较为广泛;除此之外,也有许多其他方法.小波变换也是一种被采用的方法,但此方法很难选择出合适的小波波基[19].文献[20,21]分别提出了一种鲁棒性的分段方法,实验结果也较为理想,但其主要依赖的参数范围在设定方面有一定局限性.马尔科夫建模方法[6,14]是一种基于数学模型的方法,优缺点已在上文进行了分析.

针对上述研究现状和解决方案无法处理预设信息缺失的未知步态数据的问题,本文提出了一种面向通用模型的高可用性步态周期分析方法(high feasible method of cycle segment analysis for common gait model,简称 HFM).

本文的主要贡献如下:

(1) 首先,利用快速傅里叶变换(fast Fourier transformation,简称 FFT)作为工具计算出预估步数,解决了步态信息缺失的情况下对数据进行步态周期段求解的问题;

(2) 面向步态数据分析这一领域的通用模型先验知识,设计了一种基于规则的阈值空间调整方法,避免了多种同类方法中阈值设定缺乏普适性、无法灵活应用于各个数据样本的缺点,可以更加有效且可用性更高地对于未知的步态数据进行分析,并进而求解出最准确的步态周期段;

(3) 本文还提出了一种预处理方法用以对数据进行预先过滤部分不相关数据.

本文提出的方法提高了辨别与过滤未知数据中不相关数据的能力,无须提前预知步数信息,相比以往同类型基于规则的分析识别方法,因规则的自适应性更强而可以适合更多的数据样本;并通过此方法对未知数据进行了分析并求解了步态周期数据.与实验数据集实际采样对比,取得了更为接近实际情况和后续研究需求的实验结果.

本文第 1 节简要介绍通用的步态模型及相关知识,同时详述 3 种最新具有代表性的方法思想与一种基本的通用方法,并分析其不足.第 2 节提出本文的方法并进行详述及分析.第 3 节进行详尽的实验与分析,与 3 种最新相关文献与一种经典方法进行实验对比.第 4 节对本文进行总结并对未来工作进行展望.

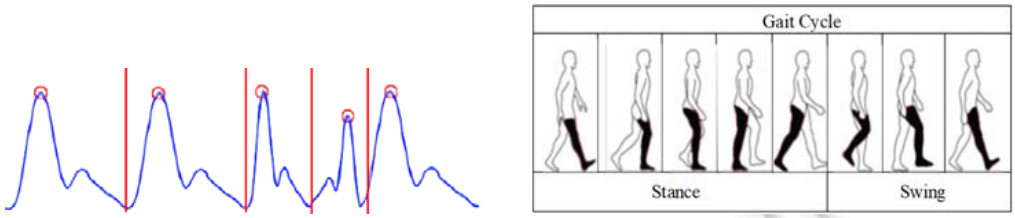
1 相关知识与对比方法

1.1 相关知识

1.1.1 通用步态模型

根据大量针对步态分析领域的文献所研究的模型可知:人类走路时,加速度计与陀螺仪主要形态都是“双峰一谷”型(如图 1(a)所示[20]),主要分为两个阶段(如图 1(b)所示[8]),其中,stance 是同侧脚跟着地到脚尖离地时身体的摆动阶段,swing 是从脚尖离地到脚跟着地时身体的摆动阶段[22].

此模型在谷点完成了两阶段的切换,基于此模型,许多文献已经做了各种步态分析方法的研发,优缺点已在第 1 节进行了说明.本文主要通过使用线加速度数据,在此通用模型的基础上,对未知的步态数据进行分析,并进一步切分为合乎通用模型的步态周期段.



(a) 步态数据通用模型[20]　　(b) 步态阶段图示[8]

Fig.1

图 1

1.1.2　快速傅里叶变换在步态时间序列周期的应用

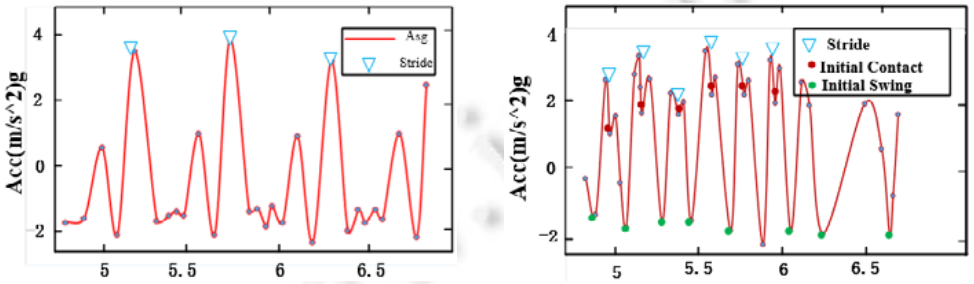
步态数据本质上是时间序列数据具备时间序列数据的各类基本特性,譬如步态的周期性与趋势性等.因此,在时间序列数据分析方面应用广泛的 FFT 与经验模态分解分析法(EMD)[23]等方法也可应用于步态数据的分析上.但根据一些资料及本文作者一定量的前期积累实验后可知,此方法需要一定的人为观察与选择,与本文的自动处理未知步态数据的目标有悖,故在此使用快速傅里叶变换的方法求解出预估的步态周期和步数.根据步态数据的实际情况,stance 与 swing 两阶段信号形态相似,基于通用模型的解决方案均认定两阶段周期相似,在本文实验中,快速傅里叶变换所取得周期为实际步态周期的 1/2.

1.2　对比方法

为了证明本文的科学性与优越性,本文对比了 3 种近年具有代表性的相关方法(基于波峰波谷的步态分析方法(OGA)[8]、基于阈值及步态模板的步态分析方法(BGR)[9]、基于标记点与阈值的步态分析方法(RRA)[5])及一种常规应用于时间序列数据周期的快速傅里叶朴素方法.

1.2.1　基于波峰波谷的步态分析方法

文献[8]提出了一种基于波峰波谷的步态分析方法,基本思想是选择比两邻大的波峰作为判定基础,将标定峰值两侧谷值作为切分点,并由此求解出步态周期段.OGA 方法的缺点在于:在实际实验中,由于各种噪声及其他问题的影响,即使进行了一定降噪与平滑处理,单纯依赖峰值分割法并不能准确切分,根据实际经验可知,只在此文图 2(a)中的理想平滑情况可行.



(a) OGA 示例[8]　　(b) OGA 误判示例

Fig.2

图 2

由图 2(a)可知,OGA 方法需要在极其平滑的数据上才能采集到该文所需的正确波峰点和波谷点.而在实际数据集中,很难控制降噪的程度与数据信息保留程度的平衡.文献[8]处理此问题方面采用 Savitzky-Golay 降噪方法[24](这是一种在信号处理及序列数据处理中广泛应用的方法,此过程包含归一化),但该文使用的自采集数据集本身噪音情况比较轻微,通过此方法可以降噪到该文方法所需平滑程度.而实际数据集(如本文的 ZJU-GaitAcc 数据集)噪声较大,在应用该文方法进行了 Savitzky-Golay 后仍然不如该文图中般平滑,若完全采用此文

方法将会出现部分噪声波峰波谷点满足条件但本身并非真实的所需点的误判情况(如图 2(b)所示).为了可以正常进行对比实验,本文对此节方法进行了略微改进(简称为 OGA+),将该文原方法中“标记的峰值点左右两侧的谷值点作为切分点”这一规则改为取比两侧均小的波谷,步骤如下.

Step 1. 降噪:对波形信号进行 Savitzky-Golay 降噪处理;

Step 2. 标记:取波形中满足比两侧波峰大这一条件的波峰点进行标记.取已标记波峰点两侧满足以下条件的谷值点并进行标记:该点小于两侧其他谷值点;

Step 3. 设定步态段首尾点:从第 1 个已标记波谷点开始,取奇数点作为步态段的起点和终点.

由后文的实验可知,此种改进(OGA+)可以较为合理地克服实际数据集噪声较大的影响,避免小谷值噪声的影响,在该文的部分实验中取得了相对好的结果,但仍不如本文主体方法进行分析所求解的周期段的符合实际情况.

1.2.2 基于阈值及步态模板的步态分析方法

文献[9]提出一种阈值及步态模板数据结合的步态分析方法,优点是比单纯依靠波峰波谷的方法更具合理性.各参数及步骤图解如图 3 所示.主要步骤如下.

Step 1. 设定参数:设定切分下限值 *stepEquilibrium*,取值为小于整体数据均值 *meanValue*,且出现次数最多的值;设定切分上限值 *stepThreshold*,取值为所有极大值从大到小排序中第 *step* 大的数值,*step* 为已知步数.

Step 2. 步态切分:若以相邻波峰(高于 *stepThreshold*)间存在低于 *stepEquilibrium* 的波谷,则说明此区间为一个步态段(如图 3 所示).

Step 3. 去除误判段:将已切分步态段与已存在模板比较,不符合模板的步态段为错误数据段.

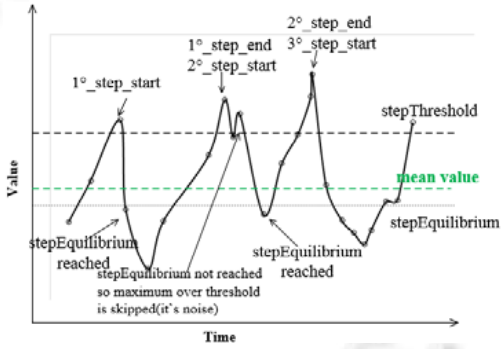


Fig.3 BGR method[9]

图 3 BGR 方法[9]

该方法的缺点是采用的步态模型与常用的步态模型均不相同,同时,必须已知步数信息与提前已计算得到的正确的步态模板数据才能对步态周期段进行分析与识别;其次,该方法是建立在频率基础上的(*stepEquilibrium* 的取值为低于 *mean value* 的值中出现频率最大的那个值),但很多数据集中,每个数据值已经唯一化,不存在最大频率值,因此,许多数据集(例如本文实验所使用 ZJUGaitAcc 数据集)无法用此方法进行分割;最后,该方法通过已知模板来剔除错误步态周期段,当没有步态模板数据或求解步态模板数据不便时,可用性受限.另外,该方法没有细分步态模型周期段内小阶段的功能.在本文使用的通用两阶段步态模型中,无法求解 stance 与 swing 阶段的位置.

1.2.3 基于标记点与阈值的步态分析方法

文献[5]是一种基于特定标记点与阈值的步态分析方法.此方法根据步态模型的实际特征定义了 4 种标记点(MS,HS,TO,Pp),通过设定这 4 种标记点的阈值,并结合通用步态周期模型对未知数据进行分析,并求解出步态周期段.主要步骤如下.

Step 1. 设定 Pp 点为波形曲线中的极大值或极小值点,定义 HS 为中间谷点,MS 为峰值点,TO 点为第 2 谷

点,作为步态切分,*zeroCrossing* 点为前后点取值分别在此值上下的阈值点.另外,该文作者根据其自采集数据集的实际情况设定了几个固定常数值 *MSThreshold*(0.5rad/s),*timeThreshold*1(350ms),*timeThreshold*2(100ms),*meanThreshold*(0rad/s),*zeroCrossing*(0rad/s);

Step 2. 当满足以下条件时可标记 MS,TO,HS,当这 3 种点均被标记完成后,一个步态段被标记完成:

Step 3. MS 点:Pp 点取值不小于 *MSThreshold*,时间位置与前一 MS 点差值不小于 *timeThreshold*1(除首个 MS 点外);

Step 4. HS 点:在已标记 MS 点的情况下,取得小于 *meanThreshold* 的 Pp 点,且满足前一 Pp 点大于 *meanThreshold*;

Step 5. TO 点:在标记 MS,HS 点的情况下,取与 HS 点位置差值不小于 *timeThreshold*2 的 Pp 点,且满足此 Pp 点前存在 *zeroCrossing* 点.

Step 6. 标记以上 3 种点后,即可完成一个步态段的切分.

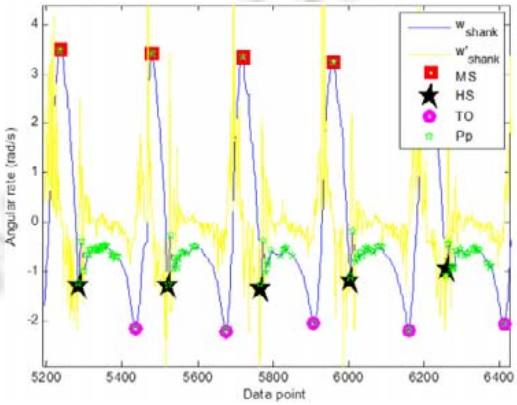


Fig.4 RRA method[5]

图 4 RRA 方法[5]

该方法的最大缺点是:阈值均为该文作者根据其自采集数据集人工设定,且是单一恒定常数,可用性受限.在实际应用中,不易总是可以人工估计到合适的阈值,在处理未知新数据时,自适应性与自修正性较差,无法灵活根据通用模型对未知新数据进行分析并切分出合理的步态周期段.

该文的原数据源为陀螺仪,由于陀螺仪与加速度计的数据均可抽象为相同的步态模型,故本文在后文对比实验中,将根据本文使用的加速度计数据集设定该文作者手动设定的几个常数阈值进行对比.

1.2.4 快速傅里叶朴素方法

FFT 也是经常应用于时间序列周期性数据的一种分析方法,步态数据是一种典型的时间序列周期性数据,许多研究也直接将这种方法应用于步态数据的分析,优点是易于使用,可以快速求解,但缺点是求解为单一固定周期数值无法适用于每步步态数据长度不同的情况,在数据集中后期,将随着数据量的增大而偏移愈发严重,同时由于此原因,使得其简单均分的通用模型内 stance 与 swing 两阶段长度没有实际意义.

针对以上 4 种方法各自的缺陷,本文提出了一种更加具有可行性的方法,根据通用模型对步态数据进行分析,从而求解步态数据周期段,并取得了较好的实验结果.

2 面向通用模型的高可用性步态周期分析

由于前文所述的目前大部分方法无法正确分析未知步态信息与未知步态模板数据、求解步态周期段时适应性较差、需要人工设定某一单一常数阈值缺乏自适应空间等缺点,本文提出了新方法来解决这些问题——面向通用模型的高可用性步态周期分析方法(a high feasible method of cycle segment analysis for common gait model,简称 HFM).

第 2.1 节首先提出一种预处理方法,其基本思想是,利用滑动窗口求解均值方法去除了首尾部分不相关数据.第 2.2 节为核心方法,即提出一种基于通用步态模型的步态周期数据分析方法.

- 第 1 阶段使用 FFT 自动地计算出预估步数及周期,后续计算将在此预估数值基础上进行.但此数值只是一个参照,并不是真实的步数及周期数值.解决了缺乏步数信息这一前提条件的问题;
- 第 2 部分在预估步数及周期的基础上,结合已有文献(如第 1.1.1 节所述)中被大量研究应用的步态模型对步态数据进行分析,改善了未知步态模板数据缺失的情况下,不易进行精确的步态周期段求解问题;同时,构建了一定自修正区间以使周期段求解的自适应性更强,改善了一些方法中因人工定义单一恒定阈值数值造成的误判问题.

第 2.3 节利用核心方法求解出的步态周期段计算出一些数据评价指标.第 2.4 节对本文方法和其他方法进行了对比与分析总结,表明了本文方法的具有更高的可用性.

2.1 步态数据预处理

本文的实验数据为 *xyz* 三轴加速度(a_x,a_y,a_z),在实际分析中通过以下公式将其转化为一维数据(a_{xi},a_{yi},a_{zi})分别为第 *i* 时刻的三轴加速度,A_{xyzi} 为该时刻转化后的总加速度):

$$|A_{xyzi}|=\sqrt{a_{xi}^2+a_{yi}^2+a_{zi}^2} \tag{1}$$

转化为一维数据后,为了去除噪声的影响,使用 Savitzky-Golay[24]方法对其进行降噪,得到脱除噪声后的数据 Asg.

在实际实验观察中,许多步态数据集的首尾都会出现不相关但也无法通过降噪方法剔除的噪声数据,本文设计了一种新的方法对这部分噪声数据进行初步去除,进而取得较为符合步态分析需求的实验数据.

Step 1. 求解数据段的整体均值 *meanData*;

Step 2. 通过使用滑动窗口从两端向内移动,滑动过程中去除窗口内最大值小于 *meanData*、最小值大于 *meanData* 的不相关数据.最终去除两端部分不需要的数据.

2.2 步态周期段求解

步态数据周期段的求解是本文的核心方法.通过根据通用模型对步态数据进行分析,进而求解得出准确的步态周期段数据.图 5 为相关参数与步骤示意图.

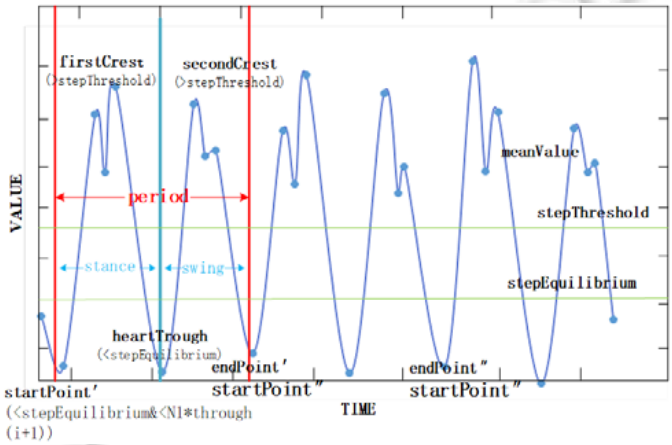


Fig.5 HFM method

图 5 HFM 方法

目前,在步态分析领域进行的研究现状主要是缺乏一种可以自适应地自动处理未知步态信息的方法,如第 1 节所示,目前此方向的研究成果具备各种缺陷,如切分规则过于简单、抗噪声信息能力差、必须已知一定量的相关信息等各种缺点.本文根据各类文献已研究出的步态数据相关知识,对已有各个方法的存在问题进行分析

与改进,得出了一种可以自适应处理未知步态信息、并可抵抗一定噪声信息影响、同时符合相关文献已研究出的步态数据模型(如第 1.1.1 节所述)的方法 HFM.HFM 方法主要分为两个阶段,步骤如下.

(1) 第 1 阶段求解预估周期和步数.

由第 1.1.2 节所述,FFT 是求解时间序列数据周期的一种基本方法,步态数据即为时间序列数据的一种,因此此处使用 FFT 方法求解出一个初始预估傅里叶变换周期,进而求解出初始平均步数(*initialStep*)与初始平均周期(*initialPeriod*).

(2) 第 2 阶段进行精确步态切分,图 5 为本部分图示.

Step 1. 参数求解:求出所有波峰、波谷及上限值 *stepThreshold* 与下限值 *stepEquilibrium*.上限值设定为此组时间序列数据的均值(*meanValue*);下限值设定为所有波谷值从小到达排序第 *initialStep* 个波谷值(如上一步 FFT 计算后的 *initialStep* 非整数则向上取整),此值必须小于均值 *meanValue*.设步态段数目为 *n*,初始值取 1;

Step 2. 标记起点:符合以下条件的波谷点($trough_i$)即为第 *i* 步步态起始点 $startPoint_i$:小于下限值,小于前谷值或后谷值的带状容错范围内(由式(2)N_1 表示,为相关文献及实验经验系数[9],此处虽是由相关文献取得的一个系数,但由于构建了一定的容错区间,与其他使用单一恒定常数的方法对比仍具有较强的普适性),同时前步需已结束标记(第 1 个步态段除外).记录下此点并初始化每个标记值:

$$\begin{aligned} trough_i &< N_1 \times trough_{i+1} \\ trough_i &< N_1 \times trough_{i-1} \end{aligned} \tag{2}$$

Step 3. 标记首波峰:当满足以下条件的波峰点即为第 *i* 步首波峰点 $firstCrest_i$:大于上限值 *stepThreshold* 且已标记过起点.记录下此点;

Step 4. 标记波谷:满足以下条件的波谷点即为第 *i* 步心波谷点 $heartTrough_i$:小于下限值 *stepEquilibrium* 且已标记过初始点及首波峰点.记录下此点,同时,此点为 stance 与 swing 阶段的分割点;

Step 5. 标记次波峰:满足以下条件的波峰点即为第 *i* 步次波峰点 $secondCrest_i$:大于上限值 *stepThreshold* 且已标记过初始点、首波峰点及心波谷点.记录下此点;

Step 6. 标记终点及下步起点:根据 Step 2 找到第 i+1 步的初始点 $startPoint_{i+1}$,并将此点作为第 *i* 步的终点 $endPoint_i$.($endPoint_i$–$startPoint_i$)为第 *i* 步步态周期 $period_i$,并通过以下条件判断 $period_i$ 是否为一个正确步态周期段:

$$period_i \in [N_2 \times initialPeriod, N_3 \times initialPeriod] \tag{3}$$

若为正确步态周期段,则步态段数目 n+1.将步态周期($period_i$)设为一个带状动态区间. $period_i$ 必须在此区间内才是一个正确的步态周期段.当 $period_i$ 取值区间在[0,(1–N_2)× *initialPeriod*]时,可能为 stance 或 swing 单一阶段,当 $period_i$ 取值区间为(N_3×*initialPeriod*, 2×*initialPeriod*]时,可能误判下一个步态周期的某一单一阶段到上一个步态周期.故根据大量实验后的经验数值与参考相关文献成果[20],给予其一定动态容错空间,将 N_2 设为 2/3,N_3 设为 4/3;

Step 7. 循环 Step 2~Step 6,直至该段数据末尾.若该段数据内最后一个数据段根据 Step 2~Step 5 计算已找到次波峰点($secondCrest_i$)则将此段数据作为一个步态段计入分析范围;否则,最末部分不计入范围;

Step 8. 结束计算,*n* 最终值为本数据集的步态段数目,并计算得出每个步态周期段的起始与终结位置.

2.3 步态数据度量

在分析数据从而得到合理的步态周期段后,可对数据进行进一步处理.由通用模型可知,人类走路主要分为以下两个阶段:stance 与 swing,本文方法和 OGA+,RRA 方法可以对步态周期段进行进一步细分.目前,由于数据信息有限,缺乏人体方向等重要信息,只通过数据集的三轴线加速度并不足以计算出行走距离、速度等信息,因此暂时只能计算数据的总体周期及两阶段周期.第 3 节实验部分对以下指标进行了计算.

(1) Estimated gait number (EGN):估计步态段数目,取值为 *n*.第 2.2 节第 2 阶段计算得出了此值;

(2) Estimated average period (EAP):估计平均步态段总周期.第 2.2 节第 2 阶段通过计算各个步态周期的 *startPoint* 与 *endPoint* 得出此值:

$$EAP = \frac{\sum_{i=1}^{n} endPoint_i - startPoint_i}{n} \tag{4}$$

(3) Estimated average stance period (EAStP):stance 阶段估计平均周期.第 2.2 节第 2 阶段通过计算各个步态周期的 *heartTrough* 与 *startPoint* 得出此值:

$$EAStP = \frac{\sum_{i=1}^{n} heartTrough_i - startPoint_i}{n} \tag{5}$$

(4) Estimated average swing period (EASwP):swing 阶段估计平均周期.第 2.2 节第 2 阶段通过计算各个步态周期的 *heartTrough* 与 *endPoint* 得出此值:

$$EASwP = \frac{\sum_{i=1}^{n} endPoint_i - heartTrough_i}{n} \tag{6}$$

(5) Period deviation (PD):平均步态段偏移比,以如下公式计算:

$$PD = \frac{|(RAP) - (EAP)|}{EAP} \tag{7}$$

其中,RAP 为 Real average period 数据集采集点的真实步态周期.

本文方法如果应用于信息更完备的数据集,可得到更多信息用于后续研究.

2.4 方法分析与总结

本文提出的方法 HFM 有效地解决了上文 4 种对比方法面临的问题.通过第 2 阶段 Step 1 对一些阈值的限定,克服了 OGA(OGA+)方法中单纯依赖峰值谷值进行切分的缺点,使得切分原则更加全面,避免了由于一些噪声影响造成的误判.

由于第 1 阶段对步数的预估与第 2 阶段各步骤对于步态模型的整体限制,无需像 BGR 方法一样必须已知步数和步态模板这两个条件,使本文方法适用范围更加广泛,可以更好地应用于后续其他研究.

在第 2 阶段 Step 6 设定了对于步态周期 *period* 的动态容错区间,避免了只用 FFT 出现的周期恒定而无法灵活适用于不同长度步态段的缺点.

同时,由于本文构建的阈值自适应区间较为合理,虽然容错系数为综合了文献与实验经验值设定而成,与 RRA 方法所设定的单一常数阈值相比仍具有更强的普适性.

综上所述,本文有效地改善了在多种信息缺失的情况下(例如步数信息与某些方法必须预置的步态模板数据)对于步态数据切分这一重要问题,同时可以在一个自适应区间内对步态数据进行正确的动态切分,进而进行进一步的数据处理,提高了方法的可用性.

3 实验与分析

3.1 实验说明

3.1.1 实验数据集

本文从 ZJU GaitAcc Dataset[10]中选择 3 条不同的数据进行实验验证,ZJU GaitAcc Dataset 是一个包含大量不同年龄、性别、身体条件情况的人步行线加速度的数据集,根据样例数据的情况,均具备步态数据通用模型的基本特征,Sample A,Sample B,Sample C 分别取自其中 3 个不同人体的手臂或手肘线加速度样本数据.

本数据集携带包含实际采样点与可用数据范围在内的两个附带信息.为了更好地进行实验,本文实验使用的数据范围是第 2.3.1 节提出的去除头尾噪声非有效步态数据方法与实际可用数据范围的交集作为实验数据的有效区间.由于实际采样点区间与数据集本身的可用数据范围并非子集关系,部分采样点是在可用数据范围之外的(这是此数据集本身已注明的情况),因而不计入方法计算.实验图中会出现部分两侧数据未被纳入实验范围的情况,在数据表格中采用实际步数栏(real gait number)“(+1)”的形式来表示可能存在在可用数据范围之

外的个别采样点个数.

该实验所使用的数据集并不一定是以波谷作为每步的起始采集点,而本文方法及对比方法均是以波谷作为起始采集点,因而可能会出现切分的平移偏差,在 Sample C 的实验中可以观察到此现象,并不完全说明本文方法或对比方法在起始点选择这一方面是错误的,Period Deviation 指标也可说明本文方法的合理性.

3.1.2 评价指标

本节实验的评价标准取为:表格中估计步态段周期 EAP、估计步态段数目 EGN 以及平均偏移量 PD.同时,也需结合各个实验图中实际采样点与各方法的步态周期切分点位置来对性能进行评价.

3.1.3 对比方法

根据 4 类不同解决此问题的思路,本文采用第 1.2 节中介绍的 4 种方法进行对比实验.

(1) 根据第 1.2.1 节的分析,OGA 本身只在理想数据下可取得该方法理论上的步态周期段求解结果,在实际多数数据集下,需进行略微改善才可正常使用,第 1.2.1 节提出了一种改善方法 OGA+.同时,对数据进行了归一化处理.

OGA(OGA+)起始点选取原则:根据该方法规则,选取本数据集有效区间内符合条件的第 1 个波谷.

(2) 根据第 1.2.2 节的分析,BGR 方法高度依赖于数据点数值的出现频率,而对数据进行降噪后,数据点取值高度趋向于唯一化而无法进行实验,因而在 BGR 方法进行实验时,采用了未降噪的原始数据(且未进行归一化),但效果仍然不佳,也说明了此方法的局限性较大.同时,方法对于错误求解的步态周期数据段采用的是与模板数据对比再剔除的方法,由于本数据集不存在模板数据,而在实际数据处理中也不一定总是存在既有标准模板数据,再加之此方法本身的切分原则与本领域对于步态模型的研究成果差别较大(见第 1.1 节介绍的通用模型),因此实验图中直接进行了标记,并未将剔除错误段的过程体现出来.此处理方法也是为了更好地呈现方法对比图.

BGR 起始点选取原则:根据该方法规则,选取本数据集有限空间内符合其理论规定的第 1 个波峰点.

(3) RRA 在本节实验中的各个常数阈值根据数据集特征分别设定为:$MSThreshold$=0.5(g),$timeThreshold1$=100(ms),$timeThreshold2$=60(ms),$meanThreshold$=0.5(g),$zeroCrossing$=0.5(g).此阈值参数适用于样本 A 的验证,同时,样本 B 与样本 C 也使用此参数验证了恒定且无容错空间的常数阈值在实际应用中的缺陷.首个 MS 点标记为 $MS1$,后续 MS 点标记为 $MS2$.同时,对数据进行了归一化处理.

RRA 起始点选取原则:由于该方法以 MS 峰值点作为算法起始点,但此方法在对步数的切分与通用模型有少量水平平移(大约小于 1/4 个步态阶段),由于本数据集多数数据样本是以波谷作为记录采样点的(个别样本采用其他点,例如 Sample C),为了更贴近数据集样本采样位置,在评价数据时,除第 1 个步态段采用第 1 个 MS 峰值点起始外,后续步态段均以 TO 波谷点作为切分.由于由于首步切除量非常小,对整体实验影响不大,同时该方法的缺陷在于恒定常数阈值问题,与此平移无关,并不影响实验证明其缺陷.

(4) FFT 的计算结果是一个恒定步态周期值,步数为有效区间总长与该周期值取商,因而计算出的结果并不一定总是整数.此方法实验仅仅为了进行对比,不作重点分析.同时,对数据进行了归一化处理.其 stance,swing 两阶段长度为简单均分.

FFT 起始点选取原则:本数据集有效区间内的第 1 个点.

(5) 本文的 HFM 方法起始点选取原则:本数据集有效区间内符合条件的第 1 个波谷.同时,对数据进行了归一化处理.

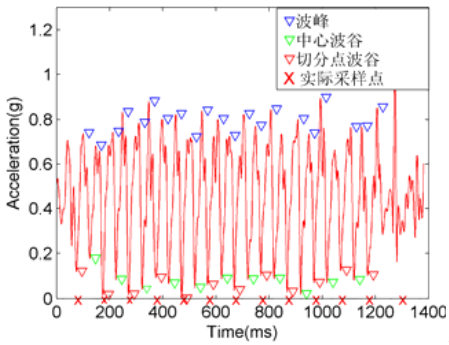
3.1.4 实验平台

本文采用 Windows 10 下 Matlab 2010b 作为实验平台,CPU 为 Intel Core P9700,内存 8GB.

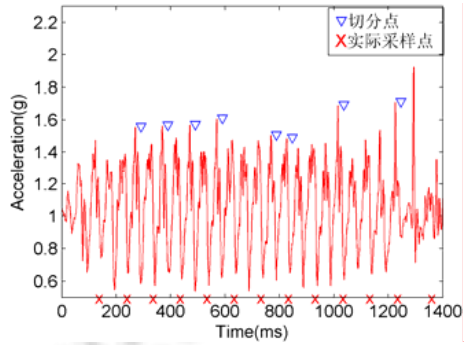
3.2 分组实验

3.2.1 Sample A

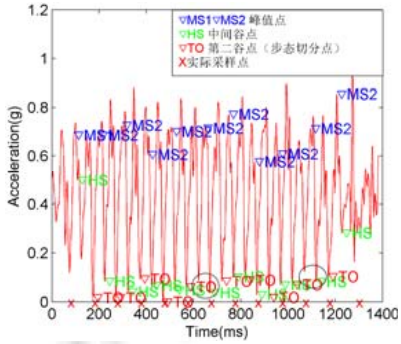
第 1 组样本中,由图 6(a)所示,各个切分点“▽”(红色倒三角)与实际采样点“X”(红色 X)均偏差不大.同时,表 1 中 OGA+的 PD 偏移量亦较小,EAP 周期接近实际值,EGN 步数在数据有效区间内也是正确的,说明本样本数据采用 OGA+方法可以取得合理效果.但与 HFM 方法相比,PD 偏移量还是略高一些.



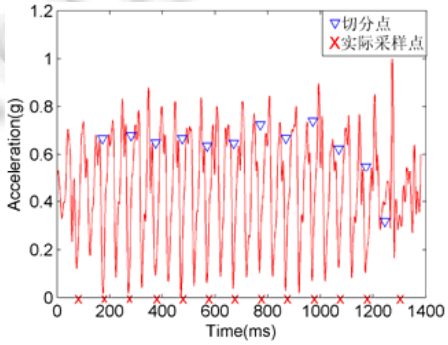
(a) OGA+



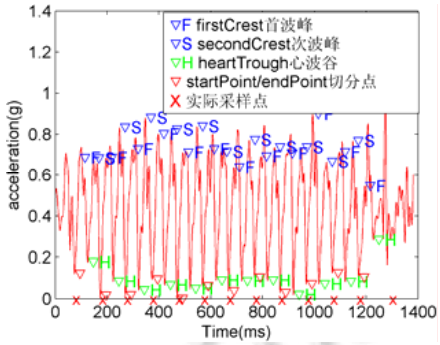
(b) BGR



(c) RRA



(d) FFT



(e) HFM

Fig.6 Signs of Sample A

图 6 Sample A 实验标记结果

Table 1 Evaluation results of Sample A

表 1 Sample A 实验数据

	OGA+	BGR	RRA	FFT	HFM
实际步态平均周期 RAP (ms)	99.73				
估计步态平均周期 EAP (ms)	99.64	136.43	107.70	99.08	99.67
平均偏移量 PD (%)	0.09	36.80	7.40	0.65	0.06
实际步数 RGN	11.00(+1.00)				
估计步数 EGN	11.00	7.00	10.00	11.99	11.00
估计 Stance 阶段平均周期 EAStP (ms)	51.08	–	56.40	49.54	51.08
估计 Swing 阶段平均周期 EASwP (ms)	49.09	–	51.30	49.54	49.75

BGR 单独用于此数据集效果不佳,由图 6(b)可见,各个切分点“▽”(红色倒三角)与实际采样点“X”(红色 X)

偏差均较大,只有极个别步态段被正确的根据理论进行了切分(例如大约 300ms~600ms 区间内的 3 段),误判率较高,PD 平均偏移量最高,EAP 周期远远大于实际值.EGN 步数与实际值相差较大.假设在有步态模板数据进行对比判断的情况下,此方法将会剔除部分不符合该步态模板数据的步态周期段.

RRA 方法在本样本中由于手动设置的阈值较为合适而表现出了可以接受的实验效果.由图 6(c)所示,前部数据各个步态周期段切分点与实际采样点偏差不大,中后部有一定偏移,较为严重的错误标记点已在圈中圈出.PD 偏移量较 BGR 方法有较大改善,但与其他方法对比仍高一个数量级.

FFT 方法计算出的 EAP 周期较为接近实际值,但由于其固定步态周期长度,导致后部数据与实际采样点偏差逐渐增大,PD 平均偏移量较 OGA+及本文 HFM 方法均大一个数量级.*EGN* 步数比实际值略大.

本文的 HFM 方法取得了最好的效果,PD 平均偏移量最小,EAP 周期较为接近实际值,且如图 6(e)所示,切分点位置较为接近实际采样点.EGN 步数在数据有效区间内也是正确的.

1 300ms 处实际采样点因已超过本数据集有效区间而没有被列入求解范围,故而 Real Gait Number 一栏填写为“11.00(+1.00)”,这是数据集本身携带的有效区间信息导致的.而非本文几种对比方法的缺陷.

3.2.2 Sample B

第 2 组样本中,OGA+方法计算出的 EAP 周期及 EGN 步数不是相差较远,根据实验图 7(a)可见,部分点按照 OGA+方法切分步态数据发生偏差,例如圆圈内(900ms 左右处)个别点标记出现错误(两红倒三角点“▽”间的绿色倒三角点“▽”为误判点),同时与图 7(a)实际采样点对比,自发生错误点(900ms 左右处)之后的数据就出现了较为严重的偏差,PD 偏移量远远大于 HFM 方法的 PD 偏移量,表明此方法在此样本中的使用性受限,而原 OGA 方法由于受到噪声波峰波谷数据影响,实验结果偏差将更大.

BGR 方法由于其本身方法的缺陷,实验结果仍然较差,只有个别步态周期段被正确切分求解,如图 7(b)所示.例如,仅有 600ms~800ms 左右处的切分区间较为符合该方法理论模型,但该方法理论模型又与常用模型均不相同,大量数据被错误标记,PD 平均偏移量、EAP 周期及 EGN 步数也远远差于其他对比方法.

RRA 方法在此样本由于其人工定义的单一阈值界限并不能灵活适应不同特征的数据集而表现不尽如人意.本样本实验采用和上一样本相同的阈值即出现了较差的实验效果.由图 7(c)可见,其步态周期段切分点与实际样本切分点与实际标记点均有较大偏移,较为严重的错误标记点已在圈中圈出,两阶段平均周期也与本文方法与 OGA+方法有着较大区别.

FFT 方法实验结果较好,如表 2 的 FFT 列,PD 平均偏差值较低,EAP 周期较为接近实际值.但由于其固定步态段长度导致了后期数据切分偏移逐渐增大,从而 PD 平均偏移量仍大于本文的 HFM 方法.EGN 步数略大于实际值.

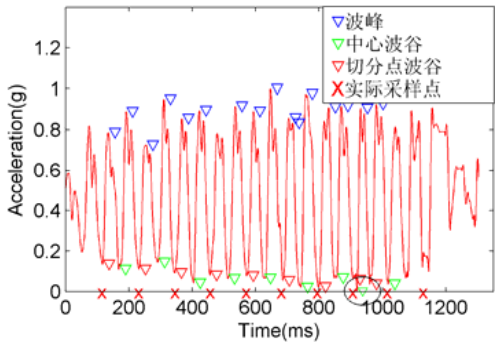
本文方法 HFM 在此数据样本中仍在各方面都取得了较为理想的效果,无论是 PD 平均偏差值或是图 7(e)中切分点与实际采样点的偏移均是所有对比方法中效果最好的.有效区间内的 EGN 步数是正确的.

本数据集在大约 1 000ms 后因不在数据集有效区间内而并未列入求解范围,故而 Real Gait Number 一栏填写为“8.00(+1.00)”.

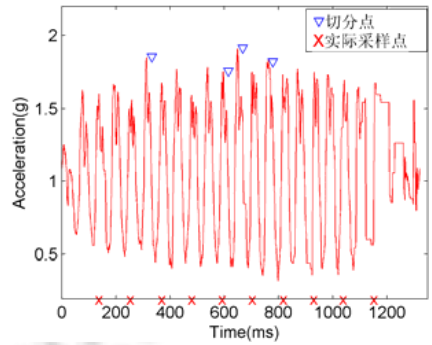
3.2.3 Sample C

本实验样本与上两个样本有一个明显的区别,由实验图实际采样点可知,本实验样本在数据采集时,实际每步的步态段起始点并非以波谷点开始,而是取了步态段数据中的其他位置,在实际采集数据时,可以以任何数据点作为步态起点.同时,由于数据集实际采样点范围与有效数据区间并不完全相同,各个方法均是在有效区间内进行,因此,本实验样本的 4 种方法在实验结果上均呈现了实际采样点与切分点的水平偏移现象,并不能说明这几种方法是错误的.关于此现象在第 3.1.1 小节已做了相关说明.

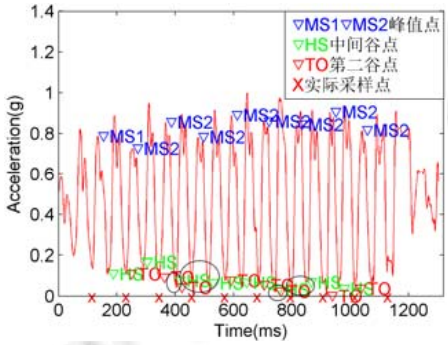
本实验样本中,OGA+方法实验图 8(a)中多个明显符合步态切分的采样点在此方法中被忽略,如 250ms 左右、400ms 左右、500ms 左右、900ms 左右处圆圈中采样点实际上是符合步态模型切分点条件的,但在此样本的 OGA+实验中没有被标记,使得最终的切分结果有误;同时,数据表 3 中,OGA+列 PD 偏移值、EAP 周期、EGN 步数也与实际值相差较大,说明了此方法在判定切分点条件方面的缺陷.



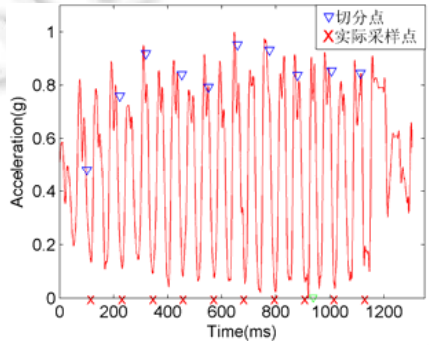
(a) OGA+



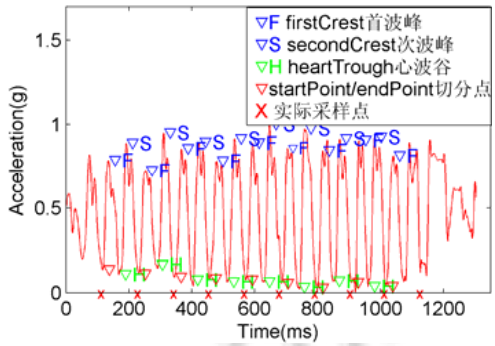
(b) BGR



(c) RRA



(d) FFT



(e) HFM

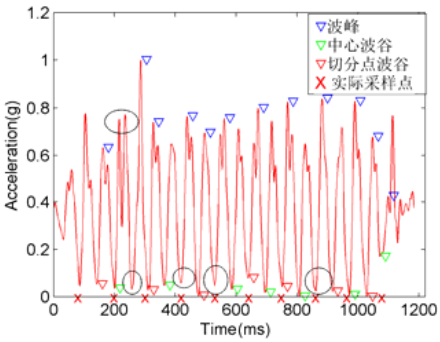
Fig.7 Signs of Sample B

图 7 Sample B 实验标记结果

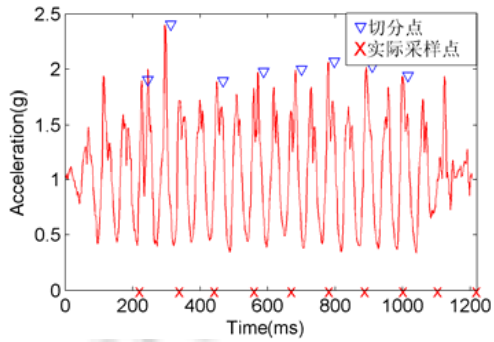
Table 2 Evaluation results of Sample B

表 2 Sample B 实验数据

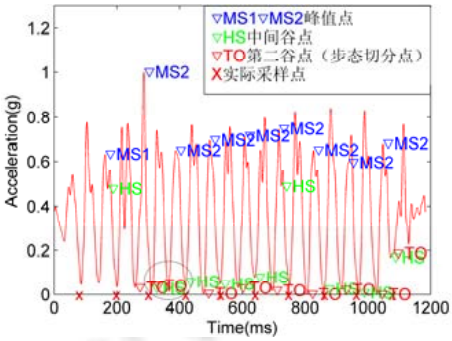
	OGA+	BGR	RRA	FFT	HFM
实际步态平均周期 RAP (ms)	112.63				
估计步态平均周期 EAP (ms)	105.75	148.67	110.37	111.83	112.78
平均偏移量 PD (%)	6.11	32.00	2.04	0.71	0.13
实际步数 RGN	8.00(+1.00)				
估计步数 EGN	8.00	3.00	8.00	8.61	8.00
估计 Stance 阶段平均周期 EAStP(ms)	51.25	–	62.50	55.92	53.13
估计 Swing 阶段平均周期 EASwP(ms)	52.63	–	47.88	55.92	59.65



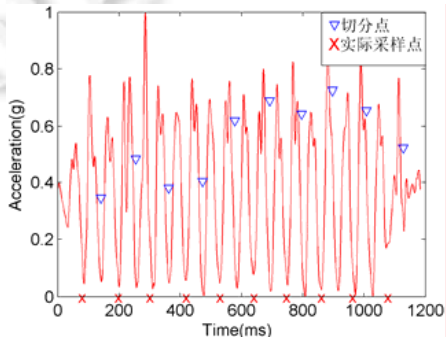
(a) OGA+



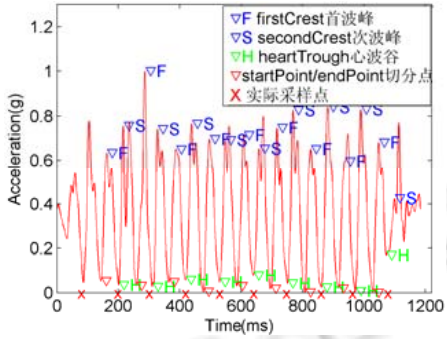
(b) BGR



(c) RRA



(d) FFT



(e) HFM

Fig.8 Signs of Sample C

图 8 Sample C 实验标记结果

Table 3 Evaluation results of Sample C

表 3 Sample C 实验数据

	OGA+	BGR	RRA	FFT	HFM
实际步态平均周期 RAP (ms)	110.89				
估计步态平均周期 EAP (ms)	147.83	110.14	102.00	112.08	110.88
平均偏移量 PD (%)	33.31	0.68	8.71	0.85	0.09
实际步数 RGN	8.00(+1.00)				
估计步数 EGN	6.00	7.00	9.00	8.74	8.00
估计 Stance 阶段平均周期 EAStP(ms)	65.00	–	50.67	56.04	54.88
估计 Swing 阶段平均周期 EASwP(ms)	82.83	–	51.33	56.04	56.00

BGR 方法在本实验样本中的表现有所好转,PD 平均偏移量较低,EAP 周期与 EGN 步数较为接近实际值.

根据图 8(b)所示,300ms~900ms 区间内的几个数据段均与实际采样点较为接近,且符合其理论.但仍有部分数据段被误判,例如 250ms~450ms 处的两段.

由于固定的人工定义单一常数阈值,RRA 方法在本样本中的效果与实际情况已经产生巨大差异.由图 8(c)所示:400ms 及 700ms 位置中 TO 切分点与 HS 中间点已经几乎贴近到同一位置,而各个切分点与实际样本切分点的偏移也较大,PD 偏移量与本文方法及 BGR,FFT 方法相差也较大.

FFT 方法在本实验样本中 PD 平均偏移量较低,但根据表 3 对比,其 PD 偏移量与本文的 HFM 方法相比仍高了一个数量级,EAP 周期较为接近实际值,EGN 步数略高.

本文的 HFM 方法仍表现最好,PD 偏移量为 4 种方法中最低,EAP 周期极为接近实际值,EGN 步数在数据有效范围内是正确的,图 8(e)中的切分结果也较为符合步态理论模型的情况.由于本数据样本实际采样点与本文方法选取起始点原则不同,出现了一定的平移偏差,在本节前文已论述过此问题,平移偏差并不能说明本部分各个方法是不合理的.

本样本在大约 1 150ms 后数据因不在有效区间内而未被列入分析范围,故而 Real Gait Number 一栏填写为“8.00(+1.00)”.

3.3 实验总结

本文实验部分总共进行了 3 组实验数据共 12 个分实验,分别验证了 OGA+,BGR,FFT 及本文改进方法 HFM 的实验效果.其中,OGA+是在 OGA 这一朴素方法的基础上进行轻微改进,以适应实际数据集的方法.

本文所使用的 ZJU-Gait Dataset 数据集自带步数及有效数据区间,本文为了体现本文方法在缺乏多种预设条件下的优越性,未使用步数信息,同时也不预设步态模板数据库.

OGA 方法及其改进 OGA+方法均为单纯基于波峰波谷的步态数据分析方法,在 3 组实验中,OGA+无论是实验对比图或是表格量化数值不如本文改进方法效果科学.原因是只依赖于峰值谷值这一单一指标,很难避免非真实步态段转换点的波峰或波谷对于步态数据段求解的影响.本文通过更为全面的求解原则,将波峰波谷检测与相关阈值结合起来,改善了这一问题.

BGR 方法的效果在本文实验中表现最差,原因在于:本文设定实验条件为位置步态步数及未预设步态模板的实验数据,BGR 方法完全无法解决这类情况;同时,该方法使用的步态模型与常用的步态模型均不相同,根据实验结果及分析可知,大量数据段被错误判断.本文通过使用 FFT 方法求解预估单独周期及步数,并在此基础上结合相关理论知识及模型进行进一步精细计算,大大克服了 BGR 方法的局限性.

RRA 方法在 Sample A 中表现较好,而在 Sample B,Sample C 中的实验效果较差.其根本原因是由于此方法建立在人工观察设定的单一恒定常数阈值上,样本 A 由于手工选定了合适的常数阈值,求解的步态周期段较为准确,Sample B,Sample C 在不改变此常数阈值取值的情况下则表现出该方法的局限性.在实际应用中,很多情况下无法事先得知或正确判断出某一数据集应适合的阈值,此方法只采用单一常数的思路无法适应这些情况,缺乏鲁棒性,适应性较低.

FFT 方法求解的结果是固定值,而实际步态数据中的步态段长度是动态变化的.因此,若使用本方法机械求解,前几步步态段可能是与实际情况相似的,但中后部的实际采样点与理论切分点的偏移必然会逐渐增大,以致于后部结果完全失去步态数据图像的规律性,为后续分析带来困难.本文通过设定周期的动态容错区间,克服了 FFT 恒定周期的这一局限.

BGR 与 FFT 方法的另一个共同缺点是:由于其方法本身有各自的功能缺陷,当需要分析步态的两个阶段时,其均无法对其进行 stance,swing 阶段切分,遑论进一步分析了.本文的解决方案中包含对于步态模型两阶段的切分过程.

文献[11]中步态段检验这一关键步骤与 RRA 方法具有相同的缺陷,同样是以一个人工定义的恒定常数阈值作为基础,缺乏可容错性空间.关于人工定义阈值的缺陷已在 BGR 及 RRA 方法的实验中得到充分体现,故不再针对此文献与本文研究课题相似的阶段再进行实验.

由表 4 的实验数据总表及上述实验图可知:本文方法的实验效果较为显著,样本 A、样本 C 均优越了一两

个数量级.尤其是从图中可见:本文方法可以较为准确地标记识别出各个步态段,而其他对比方法则出现了各种误差.实验数据中部分对比实验的偏移量较好,但多数是因正负误差抵消所致.

Table 4 General table of experiments

表 4 实验数据总表

		OGA+	BGR	RRA	FFT	HFM
Sample A	实际步态平均周期 RAP (ms)	99.73				
	估计步态平均周期 EAP (ms)	99.64	136.43	107.70	99.08	99.67
	平均偏移量 PD (%)	0.09	36.80	7.40	0.65	0.06
	实际步数 RGN	11.00(+1.00)				
	估计步数 EGN	11.00	7.00	10.00	11.99	11.00
	估计 Stance 阶段平均周期 EAStP (ms)	51.08	–	56.40	49.54	51.08
	估计 Swing 阶段平均周期 EASwP (ms)	49.09	–	51.30	49.54	49.75
Sample B	实际步态平均周期 RAP (ms)	112.63				
	估计步态平均周期 EAP (ms)	105.75	148.67	110.37	111.83	112.78
	平均偏移量 PD (%)	6.11	32.00	2.04	0.71	0.13
	实际步数 RGN	8.00(+1.00)				
	估计步数 EGN	8.00	3.00	8.00	8.61	8.00
	估计 Stance 阶段平均周期 EAStP (ms)	51.25	–	62.50	55.92	53.13
	估计 Swing 阶段平均周期 EASwP (ms)	52.63	–	47.88	55.92	59.65
Sample C	实际步态平均周期 RAP (ms)	110.89				
	估计步态平均周期 EAP (ms)	147.83	110.14	102.00	112.08	110.88
	平均偏移量 PD (%)	33.31	0.68	8.71	0.85	0.09
	实际步数 RGN	8.00(+1.00)				
	估计步数 EGN	6.00	7.00	9.00	8.74	8.00
	估计 Stance 阶段平均周期 EAStP (ms)	65.00	–	50.67	56.04	54.88
	估计 Swing 阶段平均周期 EASwP (ms)	82.83	–	51.33	56.04	56.00

根据以上实验及实验分析可知:本文提出的方法可以有效解决未知步态信息下的步态数据分析问题,求解得出了最为合理的步态周期段,且具有自适应性与自动性,对于处理与分析步态数据这一智能数据分析问题有较为理想的实验效果,是一种基于大量研究文献通用模型下的高可用性数据处理方法.

4 结论与未来工作

步态数据分析是智能数据分析的一个重要方向,针对如何更好地分析未知步态数据进而求解出更合理的步态周期段这一本领域的核心课题,本文方法有效提高了未知步态数据的分析准确性,通过预估步数,并结合相关理论知识,有效解决了目前方法无法解决的在多种信息缺失的情况下(例如步数信息与步态模板数据)对于未知步态数据进行分析并按照通用模型进行步态周期段求解这一重要问题,同时设定可自修正自适应区间,使得本文方法具有较强的动态性和可用性.

在步态数据分析方向,在此基础上还可以进行许多其他延伸课题.未来预计使用此方法结合其他信息丰富的数据集进行深入研究,如可以结合角度、GPS、磁力计信息等对步态情况进行全面研究,或结合心跳数据对心脏及其他器官的健康状况进行研究,抑或使用聚类方法或其他深度学习方法对步态进行进一步识别与其他研究.同时,也将考虑将本文成果及延伸思想应用于模式识别与数据挖掘领域.本文采用的是一种先验知识取得的模板,而在其他动作识别领域,也可以将本文思想扩展应用,通过一定方法取得新的未知动作的模板进行建模作为新的先验知识,进而对动作进行识别与分析处理,更可以将此思路应用于更广阔的智能数据处理领域.

References:

[1] Wahab Y, Bakar NA. Gait analysis measurement for sport application based on ultrasonic system. In: Proc. of the ISCE 2011. 2011. 20–24. [doi: 10.1109/ISCE.2011.5973775]

[2] Steultjens MPM, Dekker J, v. Baar ME, *et al*. Range of joint motion and disability in patients with osteoarthritis of the knee or hip. Rheumatology, 2000,39(9):955–961. [doi: 10.1093/rheumatology/39.9.955]

[3] Bora NM, Molke GV, Munot HR. Understanding human gait: A survey of traits for biometrics and biomedical applications. In: Proc. of the ICESA 2015. 2015. 723−728. [doi: 10.1109/ICESA.2015.7503444]

[4] Yuan JD, Wang ZH. Review of time series representation and classification techniques. Computer Science, 2015,42(3):1−7 (in Chinese with English abstract). http://www.jsjkx.com/jsjkx/ch/reader/view_abstract.aspx?file_no=20150301 [doi: 10.11896/j.issn.1002-137X.2015.3.001]

[5] Gouwanda D, Gopalai AA. A robust real-time gait event detection using wireless gyroscope and its application on normal and altered gaits. Medical Engineering and Physics, 2015,37(2):219−225. [doi: 10.1016/j.medengphy.2014.12.004]

[6] Taborri J, Scalona E, Palermo E, Rossi S, Cappa P. Validation of Inter-subject training for Hidden Markov models applied to gait phase detection in children with cerebral palsy. Sensors, 2015,15(9):24514−24529. [doi: 10.3390/s150924514]

[7] Formento PC, Acevedo R, Ghoussayni S, Ewins D. Gait event detection during stair walking using a rate gyroscope. Sensors, 2014, 14:5470−5485. [doi: 10.3390/s140305470]

[8] Anwary AR, Yu HN, Vassallo M. Optimal foot location for placing wearable IMU sensors and automatic feature extraction for gait analysis. IEEE Sensors Journal, 2017. [doi: 10.1109/JSEN.2017.2786587]

[9] De Marsico M, Mecca A. Biometric walk recognizer gait recognition by a single smartphone accelerometer. Multimedia Tools and Applications, 2016,76(4). [doi: 10.1007/s11042-016-3654-1]

[10] Zhang YT, Pan G, Jia K, Lu ML, Wang YM, Wu ZH. Accelerometer-based gait recognition by sparse representation of signature points with clusters. IEEE Trans. on Cybernetics, 2017,45(9):1864−1875. [doi: 10.1109/TCYB.2014.2361287]

[11] Qi YB, Soh CB, Gunawan E, Low KS, Thomas R. Assessment of foot trajectory for human gait phase detection using wireless ultrasonic sensor network. IEEE Trans. on Neural Systems and Rehabilitation Engineering, 2016,24(1):88. [doi: 10.1109/TNSRE.2015.2409123]

[12] Agostini V, Balestra G, Knaflitz M. Segmentation and classification of gait cycles. IEEE Trans. on Neural Systems and Rehabilitation Engineering, 2014,22(5). [doi: 0.1109/TNSRE.2013.2291907]

[13] Mohammed S, Sam´e A, Oukhellou L, Kong K, Huo W, Amirat Y. Recognition of gait cycle phases using wearable sensors. In: Proc. of the Robotics and Autonomous Systems. 2014. [doi: http://dx.doi.org/10.1016/j.robot.2014.10.012]

[14] De Rossi SMM, Crea S, Donati M, Reberšek P, Novak D, Vitiello N, Lenzi T, Podobnik J, Munih M, Carrozza MC. Gait segmentation using bipedal foot pressure patterns. In: Proc. of the 4th IEEE RAS/EMBS Int'l Conf. on Biomedical Robotics and Biomechatronics. 2012. 361−366. [doi: 10.1109/BioRob.2012.6290278]

[15] Moe-Nilssen R, Helbostad JL. Estimation of gait cycle characteristics by trunk accelerometry. Journal of Biomechanics, 2004,37(1): 121−126. [doi: 10.1016/S0021-9290(03)00233-1]

[16] Caldas R, Mundt M, Potthast W, Buarque F, Markert B. A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. Gait & Posture, 2017,57:204−210. [doi: 10.1016/j.gaitpost.2017.06.019]

[17] Ren L, Jones RK, Howard D. Predictive modelling of human walking over a complete gait cycle. Journal of Biomechanics, 2007, 40(7):1564−1574. [doi: 10.1016/j.jbiomech.2006.07.017]

[18] Gu X. Spectral analysis for almost periodically correlated time series [MS. Thesis]. Nanjing: Nanjing University, 2015 (in Chinese with English abstract).

[19] Fu MJ, Zhuang JJ, Hou FZ, Ning XB, Zhan QB, Shao Y. Extracting human gait series based on the wavelet transform. Acta Physica Sinica, 2010,59(6):4343−4350 (in Chinese with English abstract). http://wulixb.iphy.ac.cn/CN/Y2010/V59/I6/4343

[20] Jiang S, Wang XC, Kyrarini M, Gräser A. A robust algorithm for gait cycle segmentation. In: Proc. of the EUSIPCO 2017. 2017. [doi: 10.23919/EUSIPCO.2017.8081163]

[21] Singh G, Joshi D, Anand S, Swami P. Sub region identification in gait cycle for real time implementation. In: Proc. of the 2010 Int'l Conf. on Systems in Medicine and Biology. 2010. 108−112. [doi: 10.1109/ICSMB.2010.5735354]

[22] Dasgupta H. An algorithm for stance and swing phase detection of human gait cycle. In: Proc. of the ICECS 2015. 2015. 447−450. [doi: 10.1109/ECS.2015.7124944]

[23] Jia RY, Wang R. A similarity measure algorithm for time series based on EMD. Computer Technology and Development, 2017, 27(11):71−74 (in Chinese with English abstract). http://www.xactad.org/oa/DArticle.aspx?type=view&id=201711015 [doi: 10.3969/j.issn.1673-629X.2017.11.015]

[24] Luo J, Ying K, Bai J. Savitzky—Golay smoothing and differentiation filter for even number data. Signal Processing, 2005,85(7): 1423−1434. [doi: 10.1016/j.sigpro.2005.02.002]

附中文参考文献:

[4] 原继东,王志海.时间序列的表示与分类算法综述.计算机科学,2015,42(3):1−7. http://www.jsjkx.com/jsjkx/ch/reader/view_abstract.aspx?file_no=20150301 [doi: 10.11896/j.issn.1002-137X.2015.3.001]

[18] 顾星.几乎周期性相关时间序列的频谱分析[硕士学位论文].南京:南京大学,2015.

[19] 符懋敬,庄建军,侯凤贞,宁新宝,展庆波,邵毅.基于小波变换的人体步态序列提取.物理学报,2010,59(6):4343−4350. http://wulixb.iphy.ac.cn/CN/Y2010/V59/I6/4343

[23] 贾瑞玉,王瑞.基于 EMD 的时间序列相似性度量方法.计算机技术与发展,2017,27(11):71−74. http://www.xactad.org/oa/DArticle.aspx?type=view&id=201711015 [doi: 10.3969/j.issn.1673-629X.2017.11.015]

门慧超(1990－),女,辽宁沈阳人,博士生,主要研究领域为机器学习,数据挖掘.

王波涛(1968－),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为云计算,大数据,位置服务,隐私保护.

www.jos.org.cn