

注：国家自然科学基金（NO.61771059）；高动态导航技术北京市重点实验室开放课题

摘要：针对人体运动模式识别中最优识别特征难以确定的问题，提出一种基于 Relief-F 特征加权支持向量机的运动模式识别算法。选取 MEMS 惯性传感器的加速度时域特征构成特征向量，通过 Relief-F 算法对特征向量各元素进行权重估计，构造一个最优权重特征向量，增大不同运动模式间特征向量的差异性，采用支持向量机作为分类器，可识别站立、走、跑、跳、跌倒、上楼、下楼 7 种运动模式。实验表明，所提出的算法能够准确识别多种运动模式，识别精度可达 94.1%。

关键词：扩展过滤式选择算法；支持向量机；惯性传感器；时域特征向量；运动模式。

中图分类号：TP391.9

文献标识码：A

文章编号：1006-883X(2018)09-0028-06

收稿日期：2018-04-10

基于特征选择加权支持向量机的运动模式识别

马语晗 赵辉

北京信息科技大学高动态导航技术北京市重点实验室，北京 100101

一、引言

人体运动模式识别是指对人体多种运动状态的感知过程，主要通过通过对传感器数据的分析，实现运动模式的识别^[1]。在行人导航航位推算系统中，准确的运动模式认知，有助于准确获得行为轨迹，可应用于老弱病残等群体日常看护、行人导航、消防救援等领域^[2]。

FederalInderst 等人^[3]通过腰部穿戴 IMU 单元测量惯性数据，提取标准差、幅值范围和最大值作为分类特征，采用两级分类方式，实现分类正确率为 87.56%。Thomas Moder 等人^[4]提取行人运动的最大值、均值、标准差、均方差、四分位差和频率为分类特征，采用决策树进行运动分类与识别，识别率为 89%。NatthaponPannurat 等人^[5]提取多种时域数据作为分类特征，对八种分类算法进行评估，SVM 分类方法对运动模式识别率为 91%。

但以往方法，均存在选取的特征向量信息重叠的情况，没有解决特征冗余的问题。基于这些研究，本文基于 Relief-F 算法对 IMU 提取的特征向量进行权重估计，根据 Relief-F 算法构造一个特征元素权重向量，增大不同类特征向量的差异性，作为输入进行支持向量机分类计算，识别站立、走、跑、跳、跌倒、上楼、下楼 7 种运动状态，提高了人体运动模式识别分类的精度和效率。

二、系统介绍

1、数据采集和处理

本文采用法国 SBG Systems 公司的 MEMS 惯性测量单元 Ellipse-N 进行测试，模块安装在测试人员腰部，如图 1 所示，依照右手笛卡尔坐标系，传感器的 x 轴与地面垂直，y 轴与地面平行，z 轴指向人体运动前进方向。

对传感器原始数据预处理以提取运

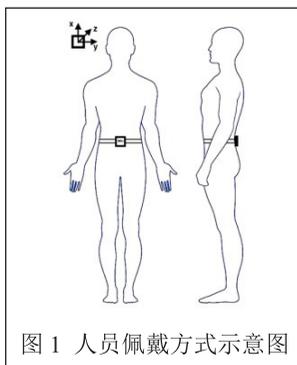


图 1 人员佩戴方式示意图

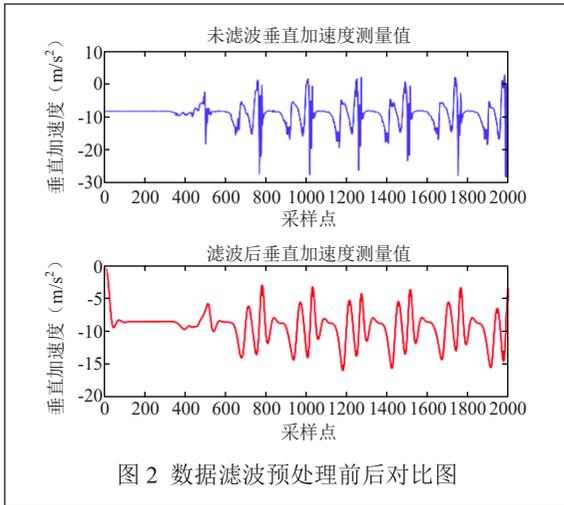


图2 数据滤波预处理前后对比图

动传感器的时域特征,采用3阶巴特沃斯低通滤波,截止频率 $f_c=5\text{Hz}$,如图2所示。可以看出,数据经过预处理后,消除了噪声,保留了加速度信号的特征,信号更加平滑,系统的分类准确度明显提高。

2、特征提取

特征提取对于多运动模式识别的精度有很大的影响。惯性测量单元 Ellipse-N 采集三轴加速度信号和三轴角速度信号,其中每轴数据都可以提取出不同的特征量。因为时域特征的提取计算复杂度低、耗时长^[6],所以本文选取适用于实时性要求较高的系统的多种时域特征。用 n 来表示一个时间窗口的大小,即窗口内数据的行数,用 i 来表示第 i 行数据。

取三轴合加速度 A 为特征提取与分类的初始数据, a_x 、 a_y 、 a_z 分别为加速度计的三轴数据。在确保精度的同时,要减少计算复杂性。合加速度公式为:

$$A = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad (1)$$

特征向量 β 由多个时域特征元素 $\beta_i (i \in \{1, 2, 3, 4, 5, 6\})$ 构成,特征向量公式为:

$$\beta = [\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6] \quad (2)$$

式中, β_1 —方差;

β_2 —四分位间距;

β_3 —峰值;

β_4 —加速度均值;

β_5 —过均值点个数;

β_6 —幅值。

(1) 方差

设 x_1, x_2, \dots, x_n 是角速度传感器数据 x 的一个样本,特征元素方差 β_1 公式如下式所示:

$$\beta_1 = \text{Variance} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3)$$

图3为多种模式下方差的分布。图中定义每50个数据点划分为一个窗口,模拟实时情况下的数据处理。

(2) 四分位间距

取特征元素 β_2 为四分位间距,把加速度计 x 轴的数据 a_x 从小到大排列为 $a_{x1}, a_{x2}, \dots, a_{xm}$,并均分为四份,处于三个分割点位置的数值即为四分位数(Quartile)。第三个四分位数和第一个四分位数的差值即为四分位间距(InterQuartile Range, IQR),计算公式为:

$$\beta_2 = \text{IQR} = Q_3 - Q_1 \quad (4)$$

式中: IQR —四分位间距;

Q_1 —第一个四分位数;

Q_3 —第三个四分位数。

四分位间距的分布情况如图4所示。

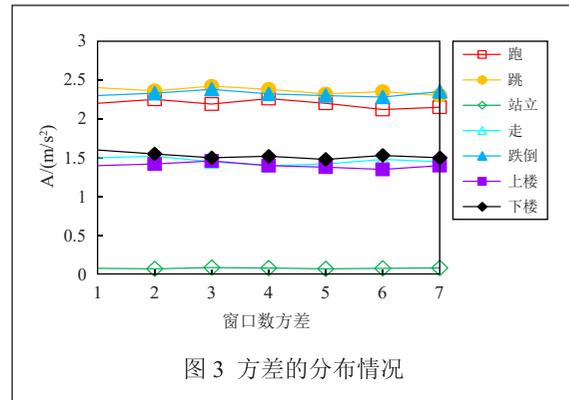


图3 方差的分布情况

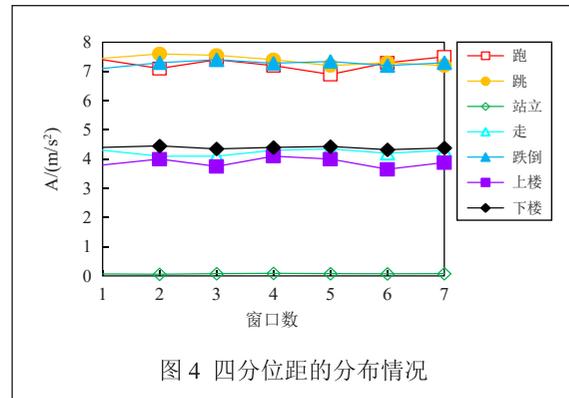


图4 四分位距的分布情况

(3) 均值

特征元素加速度均值 β_3 ，分布情况如图 5 所示，计算公式如下：

$$\beta_3 = Mean = \sum_{i=1}^n \frac{a_i}{n} \quad (5)$$

(4) 最大值

特征元素加速度最大值 β_4 ，可反应加速度信号在一个周期内的变化强度，峰值绝对值越大，说明运动幅度越大，可用来区分运动和静止的状态模式。最大值的分布情况如图 6 所示，计算公式如下。

$$\beta_4 = Maximum = |A|_{\max} \quad (6)$$

其中， A —特征元素幅值。

(5) 过均值点个数

特征元素过均值点个数 β_5 ，计算公式为：

$$\beta_5 = above_{mean} = \sum_{i=1}^n M_i \begin{cases} M_i = 1 & |A_i| > mean \\ M_i = 0 & |A_i| \leq mean \end{cases} \quad (7)$$

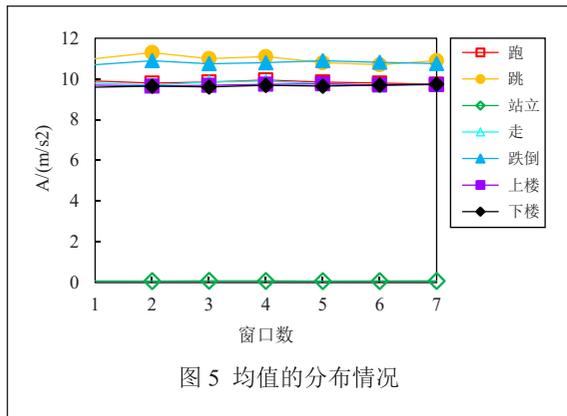


图 5 均值的分布情况

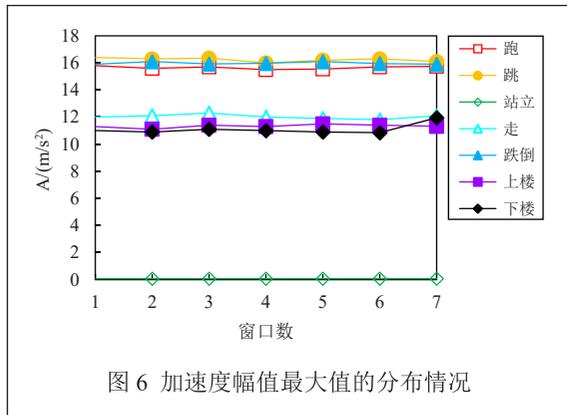


图 6 加速度幅值最大值的分布情况

(6) 幅值

特征元素幅值 β_6 ，定义为最大值和最小值之差，计算公式为：

$$\beta_6 = \Delta_{\max-\min} = |A|_{\max} - |A|_{\min} \quad (8)$$

三、算法描述^[10]

1、Relief-F 算法

Relief-F (Relevant Features) 是 Relief 算法的扩展，由 Kononenko^[7] 于 1994 年在 Kira 工作的基础上提出的一种处理多类别特征选择算法。这是一种过滤式特征选择方法，也叫特征权重算法，根据各个特征和类别的相关性赋予不同的权重，权重值小于某个阈值的特征将被剔除，仅能处理两类别数据。

设 $D = \{X_1, X_2, \dots, X_m\} (i=1, 2, \dots, m)$ 是待进行分类的对象全体， X_i 为其中任意一个样本， $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T (p=1, 2, \dots, n)$ 表示 X_i 的 n 个特征值， $\lambda_i = [a_{i1}, a_{i2}, \dots, a_{im}]^T (p=1, 2, \dots, n)$ 表示 X_i 各维特征值的权值。从 D 中随机选择样本 X_i 以后，从和 X_i 同类的样本中寻找 k 个最近邻样本 $H = \{H_1, H_2, \dots, H_k\} (j=1, 2, \dots, k)$ (特征值 $h_j = [h_{j1}, h_{j2}, \dots, h_{jn}]^T p=1, 2, \dots, n$)，称为 Near Hit，从和 X_i 不同类的样本中寻找 k 个最近邻样本 $M = \{M_1, M_2, \dots, M_k\} (j=1, 2, \dots, k)$ (特征值 $m_j = [m_{j1}, m_{j2}, \dots, m_{jn}]^T p=1, 2, \dots, n$)，称为 Near Misses。

设 $diff_hit_i$ 是 $n \times 1$ 的矩阵，表示 X_i 和 H 在每个特征上的差异：

$$diff_hit_i = \sum_{j=1}^k \frac{|x_i - h_j|}{\max(x_i) - \min(x_i)} \quad (9)$$

其中， $\max(x_i)$ 、 $\min(x_i)$ —特征值中最大值和最小值。

设 $diff_miss_i$ 是 $n \times 1$ 的矩阵，表示 X_i 和 M 在每个特征上的差异：

$$diff_miss_i = \sum_{L \in class(X_i)} \frac{P(L)}{1 - P(class(X_i))} \sum_{j=1}^k \frac{|x_i - m_j|}{\max(x_i) - \min(x_i)} \quad (10)$$

其中， $P(L)$ —待分类类别的概率，是第 L 类的样本数与数据集中样本的总数之比。

此过程重复 m 次进行迭代，权重值由下式更新：

$$\lambda_i = \lambda_i - \frac{diff_hit_i}{km} + \frac{diff_miss_i}{km}, \quad i = 1 \text{ to } m \quad (11)$$

由此，得到特征集中每一个特征的权重值向量 λ_m 。

2、SVM 分类方法

支持向量机 (Support Vector Machine, SVM) 通过构造一个分类函数或分类器, 把数据库中的数据项非线性映射到一个高维特征空间中, 从而进行线性分类, 然后在新特征空间中构造最优分类面, 形成样本分类, 可用于预测未知数据^[8]。SVM 算法简单, 鲁棒性好, 在过拟合、运算速度和分类精度等方面都比传统的分类算法具有明显的优势。

考虑线性可分的分类问题, 设训练样本集合为:

$$D=(x_p, y_i), x_i \in R^d, y_i \in \{-1,+1\}, i=1, 2, \dots, m \quad (12)$$

其中, 每一个 x_i 对应一个数据向量; y_i 是 x_i 对应的类标签; d 为空间维数。

设 d 维空间中线性判别函数的一般形式为:

$$f(x)=\text{sign}(w^*x+b^*) \quad (13)$$

其中, w —法向量;

b —偏移量, 超平面与原点的距离;

(w^*, b^*) —最优解。

根据 SVM 的基本思想, 如果超平面分类线性函数 $f(x)=\text{sign}(w^*x+b^*)$ 可以将训练样本准确的区分开, 并且使分类间隔最大即离超平面最近的距离其实是最远的, 则这个超平面是最优超平面^[9], 其解 (w^*, b^*) 为最优解。

对于线性不可分的线性支持向量机的学习问题, 变成如下凸二次规划问题:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \quad (14)$$

$$s.t. y_i(wx_i+b) \geq 1-\xi_i, \xi_i \geq 0, i=1,2,\dots,m$$

其中, ξ_i —松弛变量;

C —惩罚参数, $C>0$ 。

设问题 (14) 的解为 (w^*, b^*) , 于是可以得到最优分离超平面 $w^*x+b^*=0$ 及分类决策函数 $f(x)=\text{sign}(w^*x+b^*)$ 。

对于非线性可分的运动模式分类问题要引入核函数, 将训练样本映射到高维空间, 问题可转变为线性可分。核函数的加入将分类模型转变为下列公式:

$$f(x)=\text{sign}\left(\sum_{i=1}^m \alpha_i^* y_i K(x_i, x) + b^*\right) \quad (15)$$

其中, $K(x_i, x)$ —核函数。

则式 (14) 中的优化问题变成如下形式:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i \quad (16)$$

$$s.t. y_i(w \cdot K(x_i, x) + b) \geq 1 - \xi_i, \xi_i \geq 0, i=1, 2, \dots, m$$

3、基于 Relief-F 的 SVM 运动模式分类算法

对于非线性可分的运动模式分类问题要引入核函数, 把特征向量和核函数二者结合起来, 可以得到特征加权核函数, 然后再应用到非线性 SVM 中, 为得到较佳的实验效果, 选择高斯核函数作为高斯径向基函数分类器的核函数。

(1) 收集原始样本数据集 D :

$$\{(x_1, y_1), \dots, (x_p, y_p)\}, x_i \in R^d, y_i \in \{1, -1\}, i=1, \dots, m \quad (17)$$

(2) 根据 Relief-F 算法分别计算每个特征向量的权值 ($m=30$, 迭代 30 次), 得出下式^[7]:

$$\lambda_{30}=[0.192, 0.179, 0.183, 0.172, 0.105, 0.169]^T \quad (18)$$

剔除特征属性权重值最小的 β_5 , 从上述权重值可以看出, 方差、均值、四分位间距等都是分类主要的影响因素, 增大其权重值 (按照踢出的权重值取平均, 加在各个权重里), 由此构造新的特征权重向量:

$$\lambda=\lambda_{30}\beta=[0.213\beta_1, 0.200\beta_2, 0.204\beta_3, 0.193\beta_4, 0.190\beta_6] \quad (19)$$

(3) 选取高斯核函数 $K(x, x_c)$:

$$K(x, x_c) = \exp\left(-\frac{\|x-x_c\|^2}{2\sigma^2}\right) \quad (20)$$

其中, x_c —核函数中心;

σ —核函数的宽度参数, 控制函数的径向作用范围。

选取适当的参数 C , 构造并求解最优化问题:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j K(x, x_c) - \sum_{i=1}^m \alpha_i \quad (21)$$

$$s.t. \sum_{i=1}^m \alpha_i y_i = 0 \quad 0 \leq \alpha_i \leq C, i=1,2,\dots,m$$

其中, α —拉格朗日乘子。

求得最优解: $\alpha^*=(\alpha_1^*, \alpha_2^*, \dots, \alpha_m^*)^T$

(4) 选择 α^* 的一个正分量 $0 < \alpha_j^* < C$, 计算:

$$b^* = y_i - \sum_{i=1}^m \alpha_i^* y_i K(x, x_c) \quad (22)$$

由此构造决策函数:

$$f(x) = \text{sign}\left(\sum_{i=1}^m \alpha_i^* y_i K(x, x_c) + b^*\right) \quad (23)$$

四、实验结果与分析

为了验证本文所提出方法的有效性，设计了测试实验。实验人数 30（男性 20 名，女性 10 名），按要求完成站立、走、跑、跳、跌倒、上楼、下楼 7 种运动模式。在实验过程对测试人员的行为不作任何约束，跌倒实验在 4m×1m×0.2m 的海绵垫子上完成。

本文进行两组对比实验，分别采用多时域特征向量和基于 Relief-F 特征加权的多时域特征向量作为 SVM 运动模式识别的数据。数据分别在系统平台上实验验证，以分类准确度（一种运动模式被分类器判定为其所属类别的几率）作为评判标准。实验结果如表 1～表 4 所示。

表 1、表 2 为第一组实验数据，实验采用多时域特征向量作为 SVM 运动模式识别的数据。其中表 1 为 7 种运动模式的分类准确度。可以看出，由于与其它几种动作的区分度较大，站立的准确度达到 100%；而坐、跳和跌倒、上楼和下楼区分度较小，准确度较低。表 2 列出了 7 种运动模式的混淆情况，跳、跌倒之间易错误识别，上楼、下楼易误识别为走。

表 3 和表 4 列出了第二组实验的结果，实验采用基于 Relief-F 特征加权的多时域特征向量作为 SVM 运动模式识别的数据。对于较难区分的上楼和下楼，识别精度提高了 5%，跌倒的识别精度提高了 7%。

由此可以看出，采用本文提出的方法在训练时间上好于传统的 SVM 分类器。采用相同服务器进行测试，传统 SVM 训练平均时间为 3s，而本文平均为 1.5s，仅为传统方法耗时的 1/2，显示出其时间上的优势。

五、结束语

基于对 Relief-F 特征加权 SVM 的人体运动模式识别算法的详细阐述，对多种运动状态样本集进行了分类。实验结果表明，这种方法在分类准确率上优于传统的 SVM 分类器，具有计算准确度高、高效、简单的特点。

参考文献

[1] 涂斌斌, 谷丽华, 揣荣岩, 等. 基于 MEMS 加速度传感器的步态识别 [J]. 中国惯性技术学报, 2017, 25(3): 304-308.

[2] M Janidarmian, AR Fekr, K Radecka, et al. A Comprehensive Analysis on Wearable Acceleration Sensors in Human Activity Recognition[J]. Sensors, 2017, 17(3): 529.

[3] F Inderst, F Pascucci, M Santoni, 3D Pedestrian Dead Reckoning and Activity Classification Using Waist-mounted Inertial Measurement Unit[A]. In: 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)[C]. IEEE, 2015.

[4] T Moder, P Hafner, K Wisiol, et al. 3D Indoor Positioning with Pedestrian Dead Reckoning and Activity Recognition Based on Bayes Filtering[A]. In: 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN) [C]. IEEE, 2014.

[5] N Pannurat, S Thiemjarus, E Nantajeewarawat, et al. Analysis of Optimal Sensor Positions for Activity Classification and

表 1 第一组实验 7 种运动模式的分类准确度

运动模式	站立	走	跑	跳	跌倒	上楼	下楼
准确度(%)	100	93	94	88	86	85	84

表 2 第一组实验 7 种运动模式的混淆矩阵

运动模式	走	跑	跳	跌倒	上楼	下楼
走	186	2	0	1	6	5
跑	4	188	0	2	2	3
跳	1	3	88	6	1	1
跌倒	1	2	6	86	2	3
上楼	7	1	0	1	85	6
下楼	5	3	1	1	7	84

表 3 第二组实验 7 种运动模式的分类准确度

模式	站立	走	跑	跳	跌倒	上楼	下楼
准确度(%)	100	97	96	92	93	90	91

表 4 第二组实验中 7 种运动模式的混淆矩阵

运动模式	走	跑	跳	跌倒	上楼	下楼
走	194	1	0	0	2	3
跑	2	192	0	1	2	3
跳	1	2	92	4	1	0
跌倒	0	1	4	93	1	1
上楼	4	1	0	0	90	5
下楼	4	2	0	0	4	91

- Application on a Different Data Collection Scenario[J]. *Sensors*, 2017, 17(4): 774.
- [6] S Paiva. A Domain Independent Pedestrian Dead Reckoning System for Tracking and Localization[A]. In: 2015 IEEE 18th International Conference on Computational Science and Engineering (CSE)[C]. IEEE, 2015: 72-77.
- [7] I Kononenko. Estimating Attributes: Analysis and Extensions of RELIEF[A]. In: Proceedings of European Conference on Machine Learning, ECML' 94 [C]. Springer, Berlin, Heidelberg, 1994: 171-182.
- [8] O Aziz, J Klenk, L Schwickert, et al. Validation of Accuracy of SVM-based Fall Detection System using Real-world Fall and Non-fall Datasets[J]. *PLoS one*, 2017, 12(7): e0180318.
- [9] Schuldt C, Laptev I, Caputo B. Recognizing human actions: a local SVM approach[C]//Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. IEEE, 2004, 3: 32-36.
- [10] 李航. 统计学习方法 [M]. 北京: 清华大学出版社, 2012

Motion Pattern Recognition Based on Feature Selection Weighted Support Vector Machine

MA Yu-han, ZHAO Hui

(Beijing Key Laboratory of High Dynamic Navigation Technology, Beijing Information Science and Technology University, Beijing 100101, China)

Abstract: Aiming at the problem that the optimal recognition feature is difficult to determine in human motion pattern recognition, a motion pattern recognition algorithm based on Relief-F feature-weighted Support Vector Machine (SVM) is proposed in this paper. The time domain features of MEMS inertial sensor are selected to constitute the eigenvectors. Relief-F algorithm is used to estimate the weights of the eigenvectors. An optimal weight eigenvector is constructed to increase the difference of eigenvectors between different motion modes. As a classifier, SVM can identify 7 kinds of sport patterns of standing, walking, running, jumping, falling, upstairs and downstairs. The experimental results show that the proposed algorithm can identify a variety of motion patterns accurately with the recognition accuracy of 94.1%.

Key words: Relief-F; Support Vector Machine (SVM); inertial sensor; time domain feature vector; motion pattern

作者简介

马语晗, 北京信息科技大学高动态导航技术北京市重点实验室, 硕士研究生, 主要研究方向为行人自主导航与定位。

通讯地址: 北京市北四环中路 35 号

邮编: 100101

邮箱: moxiaohan258@163.com

赵辉, 北京信息科技大学高动态导航技术北京市重点实验室, 博士研究生, 主要研究方向为高动态导航与控制。