

本文引用格式: 李一宽. 基于多特征融合的早期帕金森病步态特征识别方法[J]. 自动化与信息工程, 2024, 45(1): 55-60.

LI Yikuan. Gait feature recognition method for early Parkinson's disease based on multi feature fusion[J]. Automation & Information Engineering, 2024, 45(1): 55-60.

## 基于多特征融合的早期帕金森病步态特征识别方法

李一宽

(重庆交通大学机电与车辆工程学院, 重庆 400074)

**摘要:** 帕金森病是一种常见的神经退行性疾病, 目前尚无可治愈的手段, 只有通过药物治疗来缓解发病周期。因此, 针对早期帕金森病的研究具有重要意义。针对早期帕金森病步态数据较少、识别准确率较低的问题, 提出一种基于多特征融合的早期帕金森病步态特征识别方法。该方法对比传统的单一特征识别方法, 识别准确率更高, 可为早期帕金森病临床诊断提供一种辅助手段。

**关键词:** 早期帕金森病; 步态数据; 特征融合; 分类识别

中图分类号: TP391.41

文献标志码: A

文章编号: 1674-2605(2024)01-0009-06

DOI: 10.3969/j.issn.1674-2605.2024.01.009

### Gait Feature Recognition Method for Early Parkinson's Disease Based on Multi Feature Fusion

LI Yikuan

(School of Mechanical and Vehicle Engineering, Chongqing Jiaotong University, Chongqing 400074, China)

**Abstract:** Parkinson's disease is a common neurodegenerative disease, and there is currently no cure available. Only drug treatment can alleviate the onset cycle. Therefore, research on early Parkinson's disease is of great significance. Aiming at the problem of limited gait data and low recognition accuracy in early Parkinson's disease, a multi feature fusion based gait feature recognition method for early Parkinson's disease is proposed. Compared with traditional single feature recognition methods, this method has higher recognition accuracy and can provide an auxiliary method for early clinical diagnosis of Parkinson's disease.

**Keywords:** early Parkinson's disease; gait data; feature fusion; classification recognition

#### 0 引言

帕金森病 (Parkinson's disease, PD) 是一种与年龄相关的神经退行性疾病。据统计, 目前全球的 PD 患者有 700 多万人, 我国的 PD 患者近 300 万人, 约占全球 PD 患者总数的一半<sup>[1]</sup>。根据专家预测, 2030 年我国的 PD 患者将达到 500 万人<sup>[2]</sup>。

目前, 临床上针对 PD 的治疗尚无可治愈的手段。因此, 早发现早治疗, 延缓病人的发病周期显得尤为重要<sup>[3]</sup>。PD 的诊断主要采用评估量表法 (H&Y 评估量表、Webster 帕金森分级量表、UPDRS 评估量表等) 和影像学法, 这两种方法主观性较强, 对医师个人能

力要求较高。根据调查发现, PD 的误诊率达 10%~50%<sup>[4]</sup>。

为提高 PD 诊断的高效性和可靠性, 许多学者将人工智能技术应用于 PD 诊断。LITTLE 等<sup>[5]</sup>利用音高周期熵 (pitch period entropy, PPE) 的声调测量方法, 提取 10 个声调高度不相关的特征向量, 并采用核支持向量机进行分类, 实验结果表明, 该方法能够很好地区分 PD 患者和健康人群。张小恒等<sup>[6]</sup>通过对 PD 语音数据集进行稀疏编码, 并对比传统稀疏编码和卷积稀疏编码, 证明了卷积稀疏编码能提取更高层的数据特征, 可更准确地区分 PD 患者与健康人群。ZHAO

等<sup>[7]</sup>通过卷积神经网络和长短期记忆网络分别从语音和 log-mel 频谱中学习相关特征, 实现 PD 的多级分类。虽然以上方法在 PD 诊断中取得了一些成果, 但仍存在一些棘手的问题, 如病人数据涉及个人隐私和医学伦理等问题, 可采集的数据较少, 导致基于数据驱动的模式识别准确率较低<sup>[8]</sup>。

为此, 本文在有限数据的情况下, 提出一种基于多特征融合的早期帕金森病步态特征识别方法, 可从不同特征中获得更多、更深层次的信息, 提高模型的分类性能。

## 1 步态数据描述与特征提取

### 1.1 步态数据描述

本文使用来自 Physionet 数据库的 PD 步态数据集。该数据集由 3 个独立的采样小组, 分别在 3 种不同的测试环境 (正常行走 Ga, 听觉刺激行走 Ju, 跑步机上行走 Si) 中对 166 名受试者进行测试。在受试者的左、右足底各安装 8 个压力传感器 (force sensitive resistors, FSR), 以 100Hz 的采样频率记录受试者足底垂直地面的反作用力 (vertical ground reaction force, VGRF), 采集时间约为 2 min, FSR 安装位置如图 1 所示。

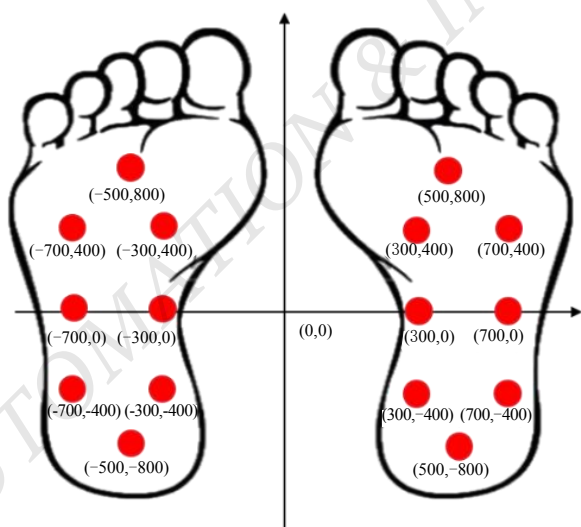


图 1 FSR 安装位置

166 名受试者中包含 73 名健康人和 93 名 PD 患者, 每名受试者的步态数据对应一个文件。该文件中

1~16 列为每个 FSR 记录的 VGRF 值随时间的变化, 17、18 列分别为左、右脚 VGRF 值的和。不同 H&Y 等级 (0 级为无症状, 等级越高 PD 程度越严重) 下相关受试者的基本物理信息如表 1 所示。

表 1 受试者的基本物理信息

H&Y	年龄/岁	身高/cm	体重/kg
0 ( $n = 73$ )	63.66±8.64	168.25±8.52	72.76±12.35
2 ( $n = 55$ )	64.22±9.62	167.96±8.31	73.18±12.19
2.5 ( $n = 28$ )	68.79±7.58	167.85±8.61	73.07±11.38
3 ( $n = 10$ )	70.80±11.23	163.11±9.91	65.56±11.21

### 1.2 步态数据预处理

受外界因素干扰, 采集的步态数据可能存在缺失值或异常值。为了使 FSR 采集的 VGRF 数据含有节律性, 先对步态数据进行预处理。考虑到受试者在测试初期需加速, 测试末期需减速, 去除 FSR 采集的前 20 s 和后 10 s 的步态数据。以 100 Hz 频率采集 2 min 的步态数据, 截取后的单个步态数据样本约含有 9 000 个样本点。考虑到受试者摆动时, FSR 可能会记录一些较小的 VGRF 值, 因此把小于 20 N 的 VGRF 值置为 0 N。《中国治疗帕金森病指南 (第三版)》<sup>[9]</sup>指出, H&Y 等级  $\leq 2.5$  为早期帕金森病患者。PD 步态数据集中 166 个数据样本筛选后, 得到健康步态数据 73 个, 早期 PD 步态数据 83 个 (55+28), 中期 PD 步态数据 10 个。

### 1.3 步态特征提取

步态是指人体步行时的姿态和行为特征, 可以反映人体的机能, 常用于临床研究, 即根据步态检测疾病<sup>[10]</sup>。人体在正常行走的过程中, 左、右脚交替轮换摆动, 具有周期性和规律性。一个完整的步态运动周期定义为同侧脚两次着地的时间差, 一个步态运动周期如图 2 所示。

一个步态运动周期分为支撑期和摆动期, 又称为支撑相和摆动相。其中, 支撑相约占整个步态运动周期的 60%, 摆动相约占整个步态运动周期的 40%。

FSR 采集的早期 PD 患者左、右脚 VGRF 时间序列如图 3 所示。

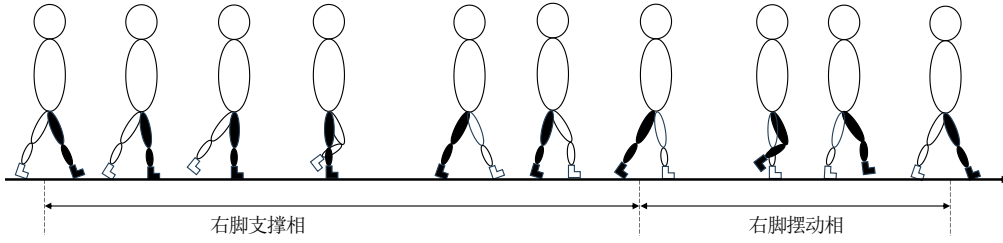


图2 一个步态运动周期示意图

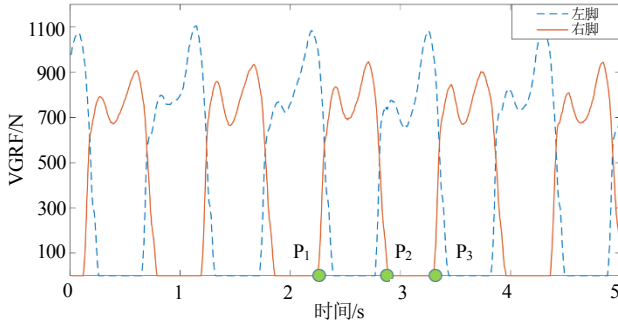


图3 早期 PD 患者左、右脚 VGRF 时间序列

图3中： $P_1$ 为右脚脚后跟开始离开地面进入单腿摆动的时间点； $P_2$ 为右脚脚尖接触地面进行单腿站立的时间点； $P_3$ 为结束站立，右脚进入下一个步态运动周期的起始时间点。

根据以上描述，以右脚 VGRF 值为例，提取相应的特征参数：站立时间、摆动时间和步幅时间等。

站立时间，单脚与地面接触的时间。在一个步态运动周期中，站立时间的起点位于 VGRF 值从 0 N 开始递增的位置，终点为 VGRF 值下降到 0 N 的位置，起点与终点间隔的样本个数为站立时间。站立时间  $S_a$  的计算公式为

$$S_a = P_2 - P_1 \quad (1)$$

摆动时间，脚与地面没有接触的时间。摆动时间内所有的 VGRF 值为 0 N，即一个步态运动周期内，第一个 VGRF 值为 0 N 到最后一个 VGRF 值为 0 N 的时间差。摆动时间  $S_w$  的计算公式为

$$S_w = P_3 - P_2 \quad (2)$$

步幅时间，站立时间和摆动时间的和。步幅时间  $S_r$  的计算公式为

$$S_r = S_a + S_w \quad (3)$$

本文利用步态特征的变异系数，体现早期 PD 患者与健康人的步态运动周期区别。步态特征的变异系数计算公式为

$$CV = \frac{M_{\text{mean}}(x)}{S_{\text{std}}(x)} \times 100\% \quad (4)$$

式中： $M_{\text{mean}}(x)$  为步态特征参数的平均值， $S_{\text{std}}(x)$  为步态特征参数的标准差。

## 2 方法

### 2.1 统计学方法

采用 IBM SPSS statistics 26 统计学软件对步态特征进行独立样本 T 检验。 $t$  值的计算公式为

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{1/n_1 + 1/n_2}} \quad (5)$$

式中： $\overline{X}_1$ 、 $\overline{X}_2$  为 2 个独立样本的平均值， $n_1$ 、 $n_2$  分别为两组样本的总数， $S_1$ 、 $S_2$  分别为样本标准差。

对两组独立样本之间差异的显著性进行统计学评估，根据  $t$  值利用统计学软件得到的结果为  $P$ 。当  $P < 0.05$  时，表明两组特征有显著差异；当  $P < 0.001$  时，表明两组特征有极其显著的差异；当  $P > 0.05$  时，无统计学意义<sup>[11]</sup>。

### 2.2 特征融合法

特征融合是一种将多个特征或特征集合结合在一起，形成一个更强大，更具有丰富信息的特征表示方法<sup>[12]</sup>，具有提高预测性能、捕捉特征的层次信息、提高模型的泛化能力等特点。考虑到以后的实验可能会提取更多的参数，本文采用特征融合法中的特征拼

接法, 首先, 将提取的特征进行归一化处理; 然后, 将特征进行向量拼接, 组合成特征矩阵; 最后, 将特征矩阵输入分类模型进行分类识别。特征拼接法可保留原始特征的信息; 帮助模型更好地捕捉特征之间的互补关系和交互影响, 从而增强模型的表达能力; 拼接不受特征类型的限制, 具有较强的模型泛化能力。

### 2.3 支持向量机算法

支持向量机 (support vector machine, SVM) 算法<sup>[13]</sup>是一种常用的监督学习方法, 主要用于分类和回归分析。SVM 算法思想是寻找一个超平面, 将不同类别的样本分开, 并且尽可能地远离所有样本, 其表达式为

$$f(x) = \omega^T \phi(x) + b = 0 \quad (6)$$

式中:  $\omega$  是超平面的法向量,  $b$  是超平面的偏置。

本文利用 SVM 算法实现早期 PD 与健康人的分类, 二分类 SVM 算法如图 4 所示。

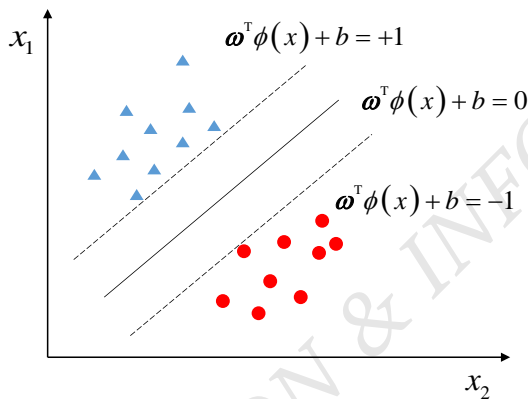


图 4 二分类 SVM 算法

在一个二分类 SVM 算法中, 当  $\omega^T \phi(x) + b = +1$  时, 则输入被分到第一类; 当  $\omega^T \phi(x) + b = -1$  时, 则输入被分到第二类。

## 3 实验结果与分析

### 3.1 步态特征分析

本文提取 73 名健康人和 83 名早期 PD 患者的左、右脚站立时间、摆动时间、步幅时间、变异系数等 12 个特征参数, 如表 2 所示。其中 CO 表示健康人群, PD 表示早期 PD 患者。

表 2 早期 PD 组与健康对照组的参数比较

步态特征参数	CO ( $n=73$ )	PD ( $n=83$ )	P 值
右脚站立时间	0.700±0.059	0.713±0.084	0.595
右脚摆动时间	0.390±0.036	0.343±0.040	<0.05
右脚步幅时间	1.090±0.087	1.056±0.111	0.314
左脚站立时间	0.722±0.072	0.730±0.078	0.749
左脚摆动时间	0.390±0.040	0.345±0.039	<0.05
左脚步幅时间	1.112±0.094	1.075±0.105	<0.05
右脚站立时间变异系数	7.418±4.408	5.024±3.057	0.067
右脚摆动时间变异系数	8.09±1.14	6.033±3.035	0.136
右脚步幅时间变异系数	9.922±6.441	7.328±5.927	0.217
左脚站立时间变异系数	6.798±3.778	4.948±3.935	0.159
左脚摆动时间变异系数	9.239±4.091	6.341±3.924	<0.05
左脚步幅时间变异系数	9.475±5.62	6.525±5.615	0.125

从表 2 可知, 早期 PD 组与健康对照组的右脚摆动时间、左脚摆动时间、左脚步幅时间和左脚摆动时间变异系数的 P 值均小于 0.05, 说明这 4 个特征参数有显著差异, 可用于区分早期 PD 患者与健康人群。

### 3.2 分类结果分析

本文将右脚摆动时间、左脚摆动时间、左脚步幅时间、左脚摆动时间变异系数以及这 4 个特征拼接的融合特征分别制作 5 种特征集, 每种类型特征集中包含健康人群和早期 PD 患者的特征参数约 14 000 个, 并将每种特征集的 80% 数据作为训练集, 20% 数据作为验证集, 输入到 SVM 模型中进行分类, 分类结果如图 5 所示。

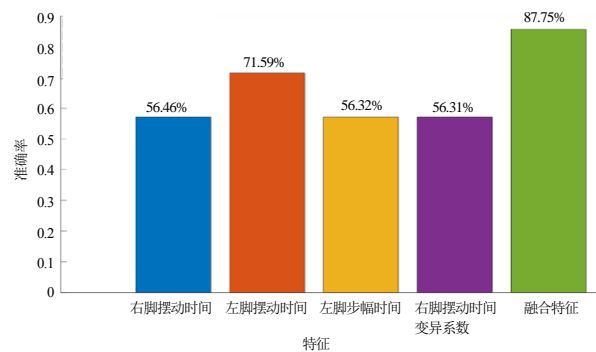


图 5 SVM 模型的分类识别准确率

由图 5 可以看出：特征拼接后的融合特征分类识别准确率最高，达到 87.75%；其次是左脚摆动时间的分类识别准确率为 71.59%，其余 3 个特征的分类识别准确率较低，表明本文方法在小样本条件下取得了较好的识别效果，相较于传统的单一特征输入方式具有一定的优越性。

为验证融合特征在不同算法中的表现，将单一特征与融合特征的特征集输入到朴素贝叶斯算法和 K 近邻算法进行分类识别，结果如图 6、7 所示。

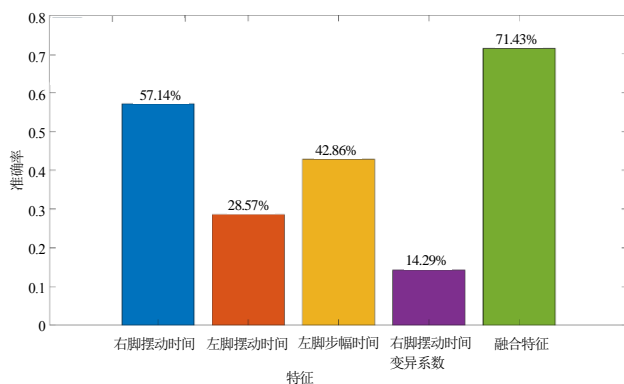


图 6 朴素贝叶斯分类识别准确率

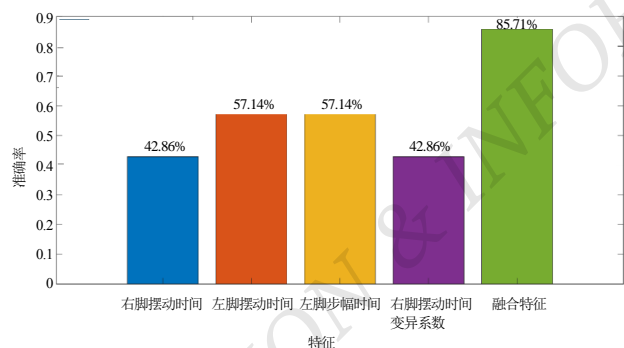


图 7 K 近邻算法分类识别准确率

由图 6、7 可知：SVM 算法对各类型特征的分类识别准确率均优于朴素贝叶斯算法和 K 近邻算法，说明 SVM 算法在小样本条件下具有较高的识别准确率；融合特征的识别准确率在 SVM 算法、朴素贝叶斯算法和 K 近邻算法中均高于单一特征的识别率，证明了本文方法的优势。

#### 4 结论

本文提出一种基于多特征融合的早期帕金森病

步态识别方法。首先，提取 VGRF 时间序列中的步态特征参数；然后，运用统计学方法寻找具有统计学意义的步态特征参数：右脚摆动时间、左脚摆动时间、左脚步幅时间、左脚摆动时间变异系数，用于区分早期 PD 患者与健康人群；最后，考虑到小样本条件下早期 PD 识别准确率较低的问题，采用多特征融合的方法将这 4 个特征融合。实验结果表明，与单一特征相比，本文方法分类识别准确率更高，可为临床医生诊断早期 PD 提供一种辅助手段。

#### 参考文献

- [1] 刘疏影,陈彪.帕金森病流行现状[J].中国现代神经疾病杂志,2016,16(2):98-101.
- [2] 韩艳,张晓红,陈彤,等.帕金森病诊治现状调查[J].中华保健医学杂志,2008,10(1):18-20.
- [3] 张杨,蒋森,刘娜,等.早期帕金森病步态障碍的临床特点及研究进展[J].医学信息,2022,35(6):34-36;40.
- [4] MEARA J, BHOWMICK BK, HOBSON P. Accuracy of diagnosis in patients with presumed parkinson's disease[J]. Age Ageing, 1999,28(2):99-102.
- [5] LITTLE M A, MCSHARRY P E, HUNTER E J, et al. Suitability of dysphonia measurements for telemonitoring of parkinson's disease[J]. IEEE Trans Biomed Eng, 2009,56(4): 1015.
- [6] 张小恒,李勇明,王品,等.基于语音卷积稀疏迁移学习和并行优选的帕金森病分类算法研究[J].电子与信息学报,2019,41(7):1641-1649.
- [7] ZHAO Jianfeng, MAO Xia, CHEN Lijiang. Speech emotion recognition using deep 1D & 2D CNN LSTM networks[J]. Biomed Signal Process Control, 2019,47:312-323.
- [8] 秦静,马雪倩,高福杰,等.基于步态分析的帕金森病辅助诊断方法综述[J].计算机应用,2023,43(6):1687-1695.
- [9] 佚名.中国帕金森病治疗指南(第三版)[J].药学与临床研究,2014,22(4):290.
- [10] 李哲.基于 EEMD 的帕金森病患者步态特征量化分级方法研究[D].长春:吉林大学,2020.
- [11] 阮桂海.数据统计与分析: SPSS 应用教程[M].北京:北京大学出版社,2005.
- [12] 高媛媛.基于多特征融合和机器学习的疾病基因检测大数据分类模型[J].微型电脑应用,2023,39(3):25-27;39.
- [13] 李福祥,王雪,张驰,等.基于边界点的支持向量机分类算法[J].陕西理工大学学报(自然科学版),2022,38(3):30-38.

作者简介:

李一宽, 男, 1996 年生, 硕士研究生, 主要研究方向: 机器人与人工智能。E-mail: 976018472@qq.com

~~~~~  
(上接第 41 页)

[16] 姚海燕,张静,留毅,等.基于多尺度小波判据和时频特征关联的电缆早期故障检测和识别方法[J].电力系统保护与控制,2015,43(9):115-123. of double-layer LSTM for 10 kV single-core cable based on multiple observable electrical quantities[J]. Electrical Engineering, 2022,104(2):603-614.

[17] CHI P, ZHANG Z, LIANG R, et al. A fault diagnosis method

作者简介:

黄艺航, 男, 1999 年生, 在读研究生, 主要研究方向: 智能测控。

蔡凯武, 男, 1984 年生, 硕士研究生, 副教授, 主要研究方向: 智能测控。E-mail: Caikw@gdut.edu.cn

黄晓智, 男, 1999 年生, 在读研究生, 主要研究方向: 智能测控。

袁澄, 男, 1999 年生, 在读研究生, 主要研究方向: 智能测控。

梁恩源, 男, 2000 年生, 在读研究生, 主要研究方向: 智能测控。

林智海, 男, 2000 年生, 在读研究生, 主要研究方向: 智能测控。

~~~~~  
(上接第 46 页)

[11] 孙玉树,李宏川,王波,等.电池储能系统 SOC 神经网络融合估计方法[J].湖南大学学报(自然科学版),2023,50(10):31-40. [12] 舒征宇,翟二杰,李镇翰,等.基于 Dropout 优化算法和 LSTM 的铅酸蓄电池容量预测[J].电源学报,2023,21(5):173-181.

作者简介:

黄晓智, 男, 1999 年生, 在读研究生, 主要研究方向: 智能测控。

张华明, 男, 1984 年生, 硕士研究生, 高级工程师, 硕士研究生导师, 主要研究方向: 智能测控。E-mail: inventor-1010@163.com

黄艺航, 男, 1999 年生, 在读研究生, 主要研究方向: 智能测控。

许志杰, 男, 1999 年生, 在读研究生, 主要研究方向: 智能测控。

~~~~~  
(上接第 54 页)

[14] 王文萍,张培恒,卫勇,基于 Kinova Mico 六自由度机械臂运动学及其路径规划的研究[J].现代机械,2022(5):58-65. [17] 周霏,陈富林,沈金龙,等.基于 MATLAB 的四自由度机械臂运动学仿真研究[J].机械制造与自动化,2016,45(1):115-119.

[15] 吴强,邓庆文,胡泽启,等.基于 RBF 神经网络与代数法的 6R 机器人逆运动求解[J].数字制造科学,2019,17(3):201-206. [18] 马强.六自由度机械臂轨迹规划研究[D].哈尔滨:哈尔滨工程大学,2007.

[16] 黄玥,张兴华,庄国锐,等.六轴焊接机械臂的逆运动学控制程序设计[J].内燃机与配件,2017(22):11. [19] 李海龙.六轴工业机械臂的结构设计与轨迹规划[D].淮南:安徽理工大学,2018.

作者简介:

姜宇凡, 女, 1998 年生, 硕士研究生, 助理工程师, 主要研究方向: 智能制造、控制工程。E-mail: 18847162723@163.com