

doi: 10.16104/j.issn.1673-1891.2023.03.009

基于改进 K-Means 聚类算法的车轮踏面损伤检测研究

朱全¹, 纪萍², 郭家伟¹

(1. 马鞍山学院智造工程学院, 安徽 马鞍山 243100; 2. 皖江工学院电气信息工程学院, 安徽 马鞍山 243031)

摘要:为提高列车车轮踏面检测效率,设计了一套基于机器视觉的车轮踏面动态检测系统,分析了 k-means 聚类算法,通过加权欧式距离对该算法进行改进,利用聚类法具有保持最大相似性的特性,将基于加权欧式距离的 k-means 聚类算法用于机器视觉的图像处理。先对原始图像作图像增强、图像灰度化等预处理,再以特征聚类思想对图像作阈值分割,使图像中的各部分特征更加突出。图像处理结果显示,基于加权欧式距离 k-means 聚类算法的车轮踏面损伤视觉检测系统可以有效地检测出踏面损伤。

关键词:车轮踏面; 动态检测; k-means 聚类算法; 机器视觉

中图分类号: U279; TP18 文献标志码: A 文章编号: 1673-1891(2023)03-0049-06

Study on Wheel Tread Damage Detection Based on Improved K-Means Clustering Algorithm

ZHU Quan¹, JI Ping², GUO Jiawei¹

(1. School of Intelligent Manufacturing Engineering, Ma'anshan University, Ma'anshan, Anhui 243100, China; 2. School of Electrical Engineering, Wanjiang University of Technology, Ma'anshan, Anhui 243031, China)

Abstract: In order to improve the efficiency of train wheel tread detection, a dynamic wheel tread detection system based on machine vision was designed, and k-means clustering algorithm was analyzed. The algorithm was improved by weighted Euclidean distance. With the clustering method's characteristics of maintaining maximum similarity, k-means clustering algorithm based on weighted Euclidean distance was used for image processing in machine vision. First, the original image was preprocessed by image enhancement and gray scale, and then the image was threshold segmented using feature clustering to make the features of each part of the image more prominent. The image processing results show that the visual detection system of wheel tread damage based on weighted Euclidean distance k-means clustering algorithm can effectively detect tread damage.

Keywords: wheel tread; dynamic detection; k-means clustering algorithm; machine vision

0 引言

随着高速铁路的快速发展,列车的安全行驶越来越受到人们的重视。车轮作为列车的关键零部件之一,其质量状况直接影响列车的安全运行。列车在行驶过程中,紧急制动时引起的轮轨之间的滑动摩擦、车轮闸瓦不结实、列车司机操作不当、车轮对轨道的冲击都会引起车轮的剥离、擦伤等踏面损伤,这些损伤会严重影响列车的安全运行和使用寿命。因此,快速、准确地检测车轮的踏面损伤,对于

列车的安全运行意义重大。目前车轮踏面损伤的检测方法,从形式上可分为静态检测和动态检测。静态检测一般在列车定期检修时由检修员通过手持工具、目测观察对车轮进行检测,这种方法耗时费力,效率较低,且存在漏检的风险。动态检测是指列车在运行状态下定性或定量检测车轮踏面损伤^[1]。常用的方法有轮轨力检测法、机器视觉检测法、声发射检测法、光纤光栅技术法、超声波检测法、轴箱振动加速度法、光纤传感器法等。其中,声发射检测法系统安装和检测都很简单,但检测信号

收稿日期: 2023-03-22

基金项目: 安徽省高校自然科学研究重点项目(KJ2021A1235)。

作者简介: 朱全(1981—),男,安徽马鞍山人,副教授,硕士,主要研究方向: 机械工程, e-mail: zhuquan612@163.com。

易受邻轮的干扰。超声波检测法的检测精度高,能适应不同车速,但系统复杂,成本较高,应用受到限制。基于图像的机器视觉检测法检测速度快、重复检测能力强,安装和使用成本较低,已经成为车轮踏面损伤检测的研究应用热点。

在基于图像的车轮踏面机器视觉检测方法中,踏面损伤图像的特征提取是一个关键环节。传统的图像分割方法如类内最大方差比法、模糊阈值分割等单阈值分割法难以准确提取踏面损伤图像的特征。因此,本文提出了一种基于加权欧式距离的k-means聚类算法的机器视觉车轮踏面损伤检测方法,用于对列车车轮踏面损伤的检测识别。

1 车轮踏面损伤类型

根据车轮损伤呈现出的不同情形,列车车轮踏面损伤分为擦伤、扁疤、剥离、凹坑磨耗、车轮多边形等,如图1所示。

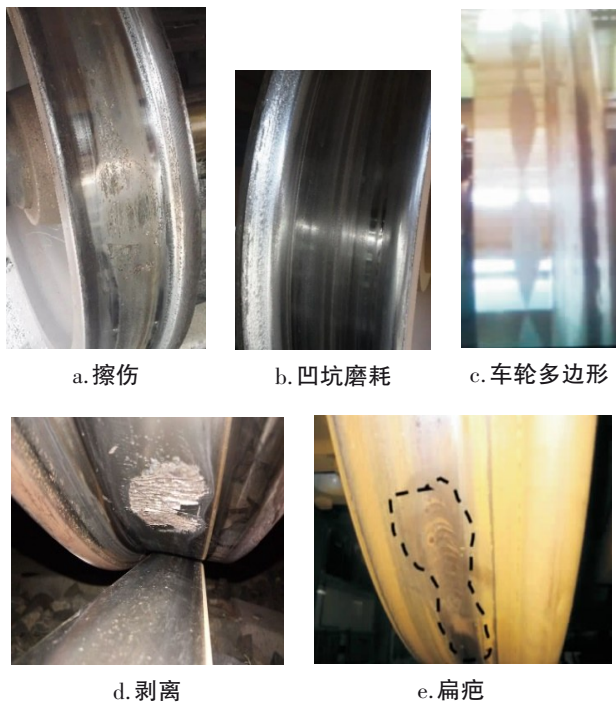


图1 车轮踏面损伤类型

擦伤是由于列车在制动过程中,车轮和钢轨之间产生剧烈摩擦,摩擦过度产生的热量使车轮踏面出现局部形变,从而形成了硬脆的马氏体组织。

凹坑磨耗是轮轨间滚动摩擦接触中发生的一种自然损耗现象。踏面凹坑对列车运行的平稳性有直接影响。从经济性角度考虑,对不同速度要求的运营列车,允许不同程度的凹坑磨耗。当凹坑磨耗到一定深度后,将致使轮轨接触不连续,影响列

车运行的安全和乘坐的舒适性。

车轮多边形是指由于轮轨磨损不均匀,车轮沿着踏面圆周方向形成的波状非均匀磨耗。车轮多边形会加大轮轨间垂向作用力,使列车的零部件容易出现疲劳失效。

剥离是指列车在正常行驶中由于轮轨接触应力或摩擦热循环作用下导致的踏面局部材料剥离掉块。车轮踏面剥离影响列车的行驶安全,减少车轮寿命,增加车轮更换的频次。

扁疤是在列车异常制动状态下,车轮运行工况改变引起轮轨间的强烈摩擦,致使车轮踏面材料部分缺失。扁疤引起的轮轨冲击载荷和动响应是正常情况下的数倍,同时产生冲击噪声,对列车其他零部件形成周期性的冲击载荷,容易引起零件的疲劳破坏。

2 车轮踏面损伤检测系统

基于机器视觉的车轮踏面损伤检测系统结构框图和总体结构图如图2、图3所示。系统由触发器、左、右两侧检测装置、交换机和上位计算机组成。左、右两侧的检测装置分别安装在轨道两侧,每侧的检测装置分别布置了1个触发器、3个光源、3台高速CCD相机。为避免发生漏检,每隔踏面周长三分之一的距离布置1台高速CCD相机。



图2 车轮踏面损伤检测系统结构

工作时,列车车轮经过触发器,高速CCD相机开始采集车轮踏面图像,通过交换机以RS485通信方式传输至上位计算机,计算机根据基于加权欧式距离的k-means聚类算法对图像做分析处理,进行踏面损伤检测,并输出检测结果。

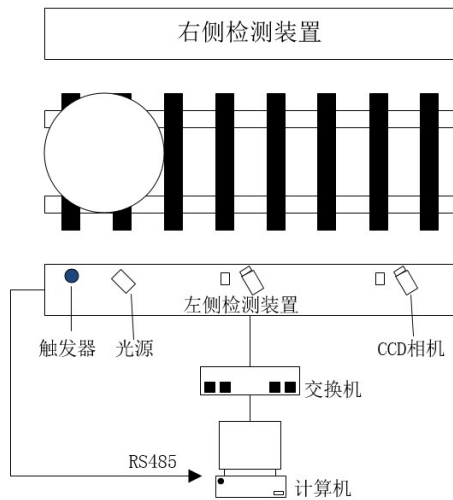


图 3 系统总体结构示意图

3 改进 k-means 聚类算法

3.1 k-means 聚类算法

聚类算法是一种无监督的学习方法,能够从分析对象的样本数据中发现关联特征,适合处理数据量大的情况,是一种强有力的数据挖掘技术。聚类算法主要包括基于划分、层次、密度、图论、网络和模型等基本算法,而 k-means 是聚类算法中一种基于划分的算法^[2]。

假设给定样本集 $\{X^{(1)}, X^{(2)}, X^{(3)}, \dots, X^{(M)}\}$, 且 $X^{(i)} \in R^N, i=1, 2, \dots, M$, 依据 Mercer 定理从映射关系 $R^N \rightarrow R^F$ 可得核函数 K , 即

$$K(X^{(i)}, X^{(i)}) = N(X^{(i)})^T N(X^{(i)}) \quad (1)$$

任何一点 $N(X^{(i)})$ 到类均值 m_k^a 的距离为

$$d_H = \|N(x) - m_k^a\| = \sqrt{K(x, x) - \frac{2}{N} \sum_{j=1}^{N_k} K(x, x_j) + \frac{1}{N_k^2} \sum_{j,p=1}^{N_k} K(x_j, x_p)} \quad (2)$$

$$m_k^a = \frac{1}{N_k} \sum_{j=1}^{N_k} N(x_j) \quad k = 1, 2, \dots, K \quad (3)$$

式中: K 为聚类数目; N_k 为第 K 类聚类的样本数目; x_j 和 x_p 则为第 K 类 C_k 的第 j 个和第 p 个样本。

根据式(2)即可求得 k-means 聚类目标函数 J 在高维核空间的表达式为

$$J = \sum_{k=1}^K \sum_{i=1}^{N_k} \|N(x) - m_k^a\|^2 \quad (4)$$

如图 4 所示, k-means 算法流程为:

- 1) 初始化聚类中心, 从数据样本中任选 k 个点作为初始聚类中心, 每个点代表一个初始聚类的中心;
- 2) 计算其他样本到各个聚类中心的欧式距离,

将该样本归到离它最近的那个聚类中心所在的类, 按下式计算欧式距离:

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (5)$$

- 3) 计算新形成的聚类的平均值得到新的聚类中心;

$$\text{meanDist} = \frac{2}{n(n-1)} \sum_{i,j=1}^n d(x_i, x_j) \quad (6)$$

- 4) 重复以上 2、3 步骤, 直到相邻两次的聚类中心没有变化, 目标函数收敛。

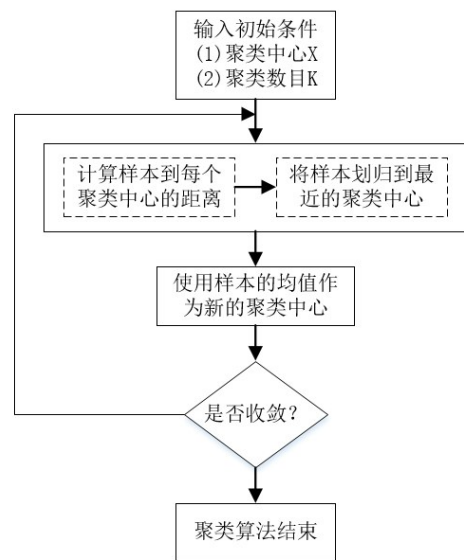


图 4 k-means 聚类算法流程图

k-means 聚类算法在每次迭代中都要考察每个样本的分类是否符合聚类条件, 若不符合, 则调整样本的分类。全部样本调整完成后, 修改聚类中心, 进行下一次迭代, 直到满足下列条件之一时迭代终止: 1) 不存在能重新分配给不同聚类的对象; 2) 聚类中心不再发生变化; 3) 误差平方和准则函数局部最小。

误差平方和准则函数定义如下:

假设给定数据集 X 包含 k 个聚类子集 X_1, X_2, \dots, X_n , 各个聚类子集中的样本数量分别为 n_1, n_2, \dots, n_k , 各个聚类子集的聚类中心分别为 $\mu_1, \mu_2, \dots, \mu_k$, 则误差平方和准则函数为:

$$E = \sum_{i=1}^k \sum_{p \in X_i} \|p - \mu_i\|^2 \quad (7)$$

3.2 改进 k-means 聚类算法

传统的 k-means 聚类算法用欧式距离作为距离度量的方法, 而欧式距离存在计算样本点与类间相似度不足的问题, 因此用相对距离取代绝对距离,

采用基于加权距离的 k-means 算法,可以更好地反映样本的真实分布,得到更合理的聚类结果。

为了反映数据样本中的数据离散度,采用变异系数赋权法确定其权值^[3]。一组数据的变异系数是它的标准差除以均值的绝对值,对数据样本的 n 个数据 x_1, x_2, \dots, x_n , 有

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (8)$$

则标准差

$$S_x = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2} \quad (9)$$

则

$$v_x = \frac{S_x}{|\bar{x}|} \quad (10)$$

式中: v_x 是 x_1, x_2, \dots, x_n 的变异系数。

对数据库中被评价对象的数据,各个属性均有各自的变异系数。用 v_i 表示 z_i 的变异系数, $i=1, 2, \dots, p$ 。属性 z_i 相应的权重系数为

$$w_i = \frac{v_i}{\sum_{i=1}^p v_i}, \quad i = 1, 2, \dots, p \quad (11)$$

改进 k-means 聚类算法的基本原理是:先计算样本数据集的各类属性的权值,再计算数据样本之间的加权欧式距离

$$d(x_i, x_j) = \sqrt{w_1(x_{i1} - x_{j1})^2 + w_2(x_{i2} - x_{j2})^2 + \dots + w_p(x_{ip} - x_{jp})^2} \quad (12)$$

改进 k-means 聚类算法流程如下:

- 1) 从数据样本集中任选 k 个点作为初始聚类中心,每个点表示聚类中心;
- 2) 根据聚类中所有点的均值,计算数据样本集中各个点与这些聚类中心的加权欧式距离,再按照最小的加权欧式距离再次对相应点进行划分;
- 3) 重新计算每个聚类的平均值,得到新的聚类中心;
- 4) 重复 2) 和 3) 步骤,进行下一次迭代,直到目标函数收敛时迭代终止,收敛条件与传统 k-means 聚类算法相同,不再赘述。

4 车轮踏面图像处理

车轮踏面损伤在图像中表现为不规则的块状图形,分布在踏面的表层。某些类型的踏面损伤区域与其他正常区域在图像中的亮度相差较大,因此,可采用区域分割的方法来实现踏面损伤的检测。由于采集的原始车轮踏面图像存在噪声信号,

需要先对图像作图像增强和灰度化,然后利用改进 k-means 聚类算法对图像进行分割,再进行二值化处理,最终确定是否存在踏面损伤,得出检测结果。处理流程如图 5 所示。

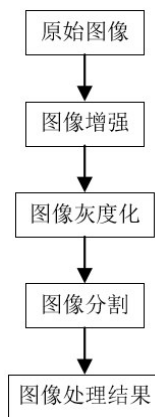


图 5 车轮踏面图像处理流程图

4.1 图像增强

图像增强的目的是使增强后的图像比原始图像更适用于特定场景的应用^[4]。经过图像增强后,目标特征更加突出,图像中的噪声信号得以抑制。图像增强方法主要为两大类:空间域增强法和频率域增强法。前者是指在空间域中,直接对图像的像素灰度值进行增强处理;后者是指在频率域下通过进行卷积计算(如傅里叶变化、小波变化等)的方法来实现图像增强处理。频率域方法中,低频主要决定图像在平滑区域中的总体灰度显示,高频决定图像细节如边缘和噪声等。为降低原始图像中噪声的影响,采用高斯低通滤波器对图像进行平滑处理,高斯低通滤波器的二维函数为

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (13)$$

式中: σ 为标准差,标准差决定信噪比大小和定位精度。

4.2 图像灰度化

采集的原始图像为 RGB 彩色模型,图像中的每个像素点都含有 R(红色)、G(绿色)、B(蓝色)三色分量,每个分量的范围为 0~255。灰度化是指将三色分量数值作等值化处理($R=G=B$),此时每个像素点只有一个字节的灰度值,范围为 0~255。图像经过灰度化处理后占用计算机内存较少,提高了计算机的处理速度。常用的灰度化处理有 3 种方法:最大值法、平均值法和加权平均法。相较最大值法和平均值法,加权平均法得到的灰度图像效果最好。加权平均法的灰度值按下式计算

$$\text{Gray} = 0.299R + 0.578G + 0.114B \quad (14)$$

将图像中每个像素计算出各点的灰度值,即可完成图像的灰度化转换。

4.3 图像分割

图像分割可以简化图像分析、处理和决策过程。根据图像的灰度、纹理或形状等参数,将其划分为不同的区域,并使这些参数在同一区域呈现相似性,而在不同区域之间则呈现明显的差异性。图像分割主要有3种方法:阈值法、区域法和边缘检测法。阈值法是图像分割中常用的方法,其中的聚类法属于全局阈值方法,利用聚类思想保持最大相似性,通过多次迭代获取最佳阈值。

采用改进k-means聚类算法对图像进行分割,将图像的每个像素点的灰度值作为样本,整个图像作为样本集合,这样就吧图像分割处理转换为数据集合的聚类任务。然后,在样本集合中运用改进k-means聚类算法进行图像区域分割,最后提取图像区域特征。根据改进k-means聚类算法原理,设置聚类数目为4类,分别代表踏面正常区域、踏面损伤区域、背景区域、其他区域。经过改进k-means聚类算法分割后的图像,损伤区域和正常区域经初步分割,两者的灰度值有明显差异,可以据此判定车轮踏面是否存在损伤。

4.4 图像处理结果

图6、7、8分别从车轮踏面损伤图像中,选取了车轮踏面擦伤、踏面剥离和车轮多边形图像。先对采集的原始图像灰度化,再分别用传统K-means聚类算法和改进K-means聚类算法进行图像分割,并比较了两种方法在图像分割方面的处理效果。以踏面擦伤为例,图6c、6d分别是采用改进K-means聚类算法和传统K-means聚类算法作的图像分割。不难看出,采用前者作的图像分割,踏面擦伤区域和正常区域的灰度值差异性比采用后者作图像分割的差异性更大,因此采用改进K-means聚类算法可以有效地检测出存在踏面擦伤的车轮。图7为踏面剥离的图像处理,结果与踏面擦伤类似。

图8为车轮多边形的图像处理结果,相较于图6、7可以发现,无论是采用传统K-means聚类算法还是采用改进K-means聚类算法,踏面损伤区域和正常区域分割的结果都不完全,两者的灰度值差异不大。究其原因主要是由于车轮多边形原始图像中踏面的损伤区域和正常区域并不明显,同时受到光源照射的影响,使得处理效果不佳。因此可以认为,本文算法适用于踏面损伤区域和正常区域存在

一定差异性的损伤。

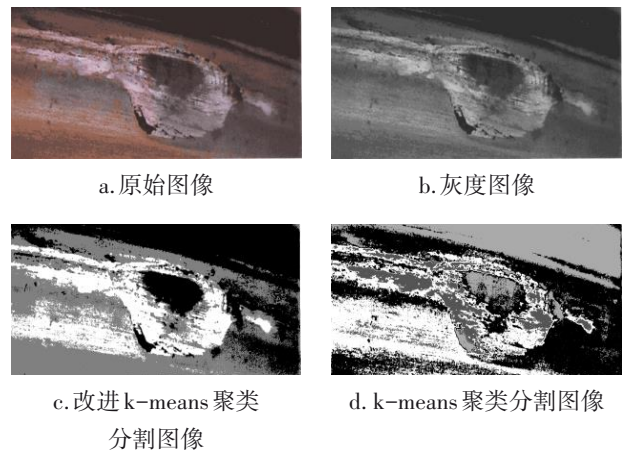


图6 擦伤图像处理结果

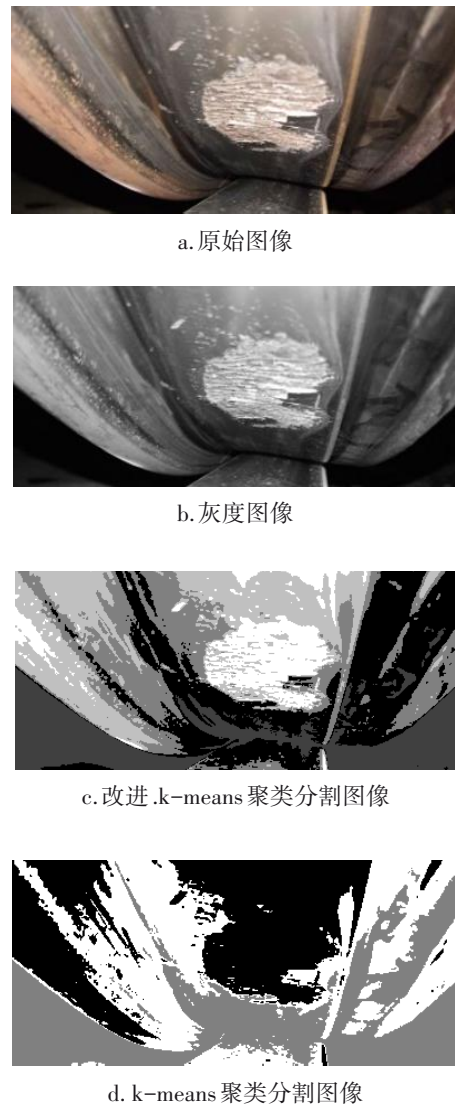


图7 剥离图像处理结果

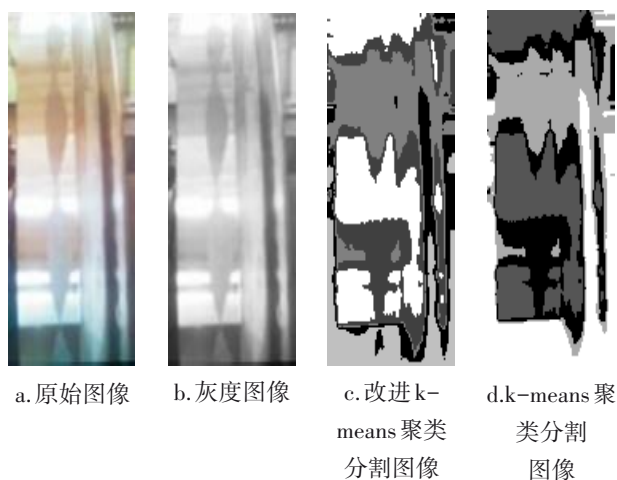


图 8 车轮多边形图像处理结果

5 结论

1) 设计了一套基于加权欧式距离的 k-means 聚类算法的车轮踏面损伤视觉检测系统,能够快速、准确地进行车轮踏面损伤检测。

2) 采用基于加权距离的 k-means 算法,可以更好地反映样本的真实分布,得到更合理的聚类结果。利用聚类法具有保持最大相似性的特性,将该算法用于图像处理,以特征聚类思想对图像作阈值分割,使图像中的各部分特征更加突出。

3) 图像处理结果显示,采用基于加权距离的 k-means 聚类算法用于车轮踏面损伤视觉检测系统,可以有效地检测出车轮踏面损伤区域与正常区域有差异性的损伤。

参考文献:

- [1] 肖乾,姜雄峰,刘海涛,等.铁道车轮踏面损伤实时监测方法研究综述[J].华东交通大学学报,2021,38(4):99-112+126.
- [2] 唐林.自动采摘目标图像快速识别算法研究——基于 k-means 聚类算法[J].农机化研究,2023,45(5):32-36.
- [3] 张忠林,曹志宇,李元韬.基于加权欧式距离的 k-means 算法研究[J].郑州大学学报(工学版),2010,31(1):89-92.
- [4] 宋志明.列车车轮踏面擦伤图像检测算法研究[D].成都:西南交通大学,2012.
- [5] 侯涛,张志腾.改进 Canny 算子在列车车轮踏面损伤检测中的应用[J].铁道科学与工程学报,2018,15(8):2107-2112.
- [6] 金学松,赵国堂,梁树林,等.高速铁路轮轨磨损特征、机理、影响和对策——车轮踏面横向磨耗[J].机械工程学报,2018,54(4):3-13.
- [7] ALEMI A, CORMAN F, LODEWIJKS G. Condition monitoring approaches for the detection of railway wheel defects [J]. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2017, 231(8):961-981.
- [8] 杨颖.基于图像的踏面表面缺陷特征分析识别[D].南京:南京航空航天大学,2017.
- [9] 曹帅帅,陈雪鑫,苗圃,等.基于 PSO 与 K-均值聚类算法优化结合的图像分割方法[J].计算机与现代化,2020(1):22-27.
- [10] 李丽亚,闫宏印.改进 k-means 加权自适应多视图数据聚类算法[J].计算机仿真,2021,38(8):314-317+429.

(上接第 48 页)

参考文献:

- [1] 周龙保.内燃机学[M].北京:机械工业出版社,2015.
- [2] 程小钢,周斌,张钊,等.低速提扭对废气涡轮增压柴油机燃烧排放特性的影响[J].汽车技术,2019(5):53-57.
- [3] 闫嘉楠,周斌,南靖雯,等.电动增压器不同布置方式对涡轮增压器的影响[J].内燃机,2018(4):5-9+13.
- [4] 刘青林,何博,滕勤.高效 Atkinson 循环 TGDI 发动机作为传统动力的研究[J].内燃机工程,2017,38(6):131-138.
- [5] 秦朝举,杨振中,张卫正,等.氢发动机低温燃烧特性的研究[J].汽车工程,2017,39(2):133-137.
- [6] 韩敬贤,黄昭明,王利,等.LP-EGR 和电增压协同技术对汽油机性能的影响[J].机械科学与技术,2022,41(1):82-87.
- [7] 丁占铭,孔冰,程振宇,等.电动增压动态响应特性试验研究[J].车用发动机,2021(6):30-36.
- [8] 黄英铭,吕伟,李云龙,等.汽油机电动增压系统研究[J].车用发动机,2021(5):87-92.
- [9] 杨攀涛,崔涛,赵彦凯,等.电动增压器对柴油发动机低速稳态性能的影响[J].兵工学报,2021,42(9):1829-1837.
- [10] 张卫波,杨南杰,吴星.车用电动增压系统的设计及其性能试验研究[J].福州大学学报(自然科学版),2017,45(6):866-872.