

考虑时间序列模糊分割的多波段激光数据分类

刘艳¹, 黄亚博²

(1. 郑州工程技术学院 信息工程学院, 河南 郑州 450044; 2. 河南大学 计算机与信息工程学院, 河南 开封 475004)

摘要: 激光扫描仪在不同波段上的测量受到大气条件、目标表面等多种因素的反射特性影响,使得数据在波段间存在交叉敏感性,导致数据在时间序列上表现出不确定性,从而降低了激光数据的分类精度。为此,提出考虑时间序列模糊分割的多波段激光数据分类。联合主成分分析方法以及遗传算法,提取多波段激光数据的特征,并找出其中最大特征矢量,构建多波段激光数据的时间序列;使用遗传算法对激光数据序列实施优化处理,减少了交叉敏感性对分类结果的影响;使用模糊分割算法将优化后的激光数据序列分割成若干时间序列段,联合 K-means 算法完成序列段聚类,实现多波段激光数据的精准分类,增加了多波段激光数据的分类精度。实验结果表明,利用该方法开展多波段激光数据分类时,分类精度高、分类效果好。

关键词: 时间序列;模糊分割;多波段激光数据;K-means 算法;主成分分析方法

中图分类号: TN914; TP391 **文献标识码:** A **DOI:** 10.3969/j.issn.1001-5078.2025.04.006

Multi-band laser data classification considering fuzzy segmentation of time series

LIU Yan¹, HUANG Ya-bo²

(1. College of Information Engineering, Zhengzhou University of Technology, Zhengzhou 450044, China;

2. College of Computer and Information Engineering, Henan University, Kaifeng 475004, China)

Abstract: The measurement of laser scanner in different bands is affected by the reflection characteristics of atmospheric conditions, target surface and other factors, which makes the data cross-sensitivity between bands, resulting in data exhibiting uncertainty in the time series, thus diminishing the classification accuracy of the laser data. For this reason, a multi-band laser data classification considering time series fuzzy segmentation is proposed. Firstly, combined with principal component analysis and genetic algorithm, the features of multi-band laser data are extracted, the maximum feature vector is found out, and the time series of multi-band laser data is constructed. Secondly, the genetic algorithm is employed to optimize the laser data sequence, which mitigates the impact of cross sensitivity on the classification results. Then, the fuzzy segmentation algorithm is utilized to divide the optimized laser data sequence into several time series segments, and in conjunction with the K-means algorithm, sequence segment clustering is completed to achieve precise classification of multi-band laser data and enhance its classification accuracy. The experimental results show that this method yields high classification accuracy and excellent classification effects when applied to multi-band laser data classification.

基金项目: 河南省高等教育教学改革研究与实践项目 (No. 2024SJGLX0570); 河南省科技攻关计划 (工业领域) 项目 (No. 212102210571) 资助。

作者简介: 刘艳 (1980-), 女, 硕士, 副教授, 研究方向为物联网、大数据。E-mail: wws02587@163.com

收稿日期: 2024-07-22; **修订日期:** 2024-09-12

Keywords: time series; fuzzy segmentation; multi-band laser data; K-means algorithm; principal component analysis method

1 引言

激光扫描仪(如激光雷达 LiDAR)已成为获取地表信息的重要工具。激光扫描仪能够在不同波段上测量,提供丰富的空间和光谱信息,对于环境监测、城市规划、农业管理等领域具有重要价值^[1-2]。不同波段的数据受到大气条件、目标表面反射特性等多种因素的影响,这些因素导致数据在波段间存在交叉敏感性,使得多波段激光数据在时间序列上表现出不确定性,导致无法有效处理这种复杂的数据特性。

文献[3]方法归一化处理获取的光谱数据,联合随机森林算法以及主成分分析设计用于特征提取的 PCA-RF 特征提取算法,从而实现数据的分类。文献[4]方法将数据输入至根据 Logistic 回归理论建立的数据分类模型中,计算后期数据的因变量以及自变量,提取数据统计特征,完成数据分类。文献[5]方法建立特征融合模型,提取激光数据的全波形时序特征以及点云几何特征,采用邻域特征增强算法挖掘数据特征,联合层次化解密器搭建分类网络,实现数据的有效分类。文献[6]方法在自觉模糊理论以及粗糙集理论下,联合 Bayesian 概率设计了一个 Bayesian 直觉模糊粗糙集模型;将数据输入模型中,获取数据的直觉模糊等价关系;根据直觉模糊集和截集的特性,实现数据的分类。

为解决上述数据分类方法中存在的问题,提出考虑时间序列模糊分割的多波段激光数据分类。

2 多波段激光数据时间序列构建

激光扫描仪在不同波段上的测量能够捕捉到目标在不同时间点的物理和化学特性变化。然而,大气中的水汽、尘埃、气溶胶等成分会吸收和散射激光,尤其是在蓝色、绿色和红色波段。这种吸收和散射作用会随着时间和天气条件的变化而变化,导致激光数据的时间序列出现波动^[7-8]。采用主成分分析方法与遗传算法,提取多波段激光数据的特征,并找出其中最大特征矢量,构建多波段激光数据的时间序列,识别和校正由于大气条件和目标表面反射特性变化引起的数据偏差,从而提高分类的准确性。

2.1 多波段激光数据时序特征提取

受到大气条件、目标表面等多种因素的反射特

性影响,导致激光数据在波段间存在交叉敏感性。通过提取时序特征,可以更好地理解不同波段数据随时间变化的规律,从而构建出多波段激光数据时间序列,减少交叉敏感性对分类结果的影响。

主成分分析方法(Principal Component Analysis, PCA)方法能够有效的去除原始多波段激光数据中的冗余数据,提取出数据中方差贡献率大的主成分因子,因此可将其应用于多波段激光数据中,对多波段激光数据展开特征提取。具体提取步骤如下:

(1)将多波段的激光数据波段设定成 a_i 形式,不同波段数据提取的信息概率设定成 h_i 形式,基于样本空间以及概率空间对多波段激光数据展开表述,结果如下式所示:

$$\begin{pmatrix} a \\ h \end{pmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ h_1 & h_2 & \cdots & h_n \end{bmatrix} \quad (1)$$

完成描述后,基于表述结果构建多波段的激光数据表述成矩阵形式 $Q_{m \times n} = a_i \times h_i$,其中数据样本数量记作 m ,激光数据的波段数量记作 n ,以此建立多波段激光数据的协方差矩阵 C :

$$C = \frac{Q_{m \times n} \times Q_{m \times n}^T}{n - 1} \quad (2)$$

(2)完成协方差矩阵建立后,对矩阵展开求解计算获取第 i 个激光数据的波段的特征值 γ_i ,并对求解结果降序排列处理。特征值 γ_i 计算为:

$$\gamma_i = C \times x_i \quad (3)$$

式中, x_i 表示单位化特征向量。

完成特征向量计算后,针对激光数据协方差矩阵特征值,获取第 i 个激光数据的波段概率 h_i ,联合香农信息理论,计算数据特征向量值包含的信息量 I_i 结果如下式所示:

$$\begin{cases} h_i = \frac{\gamma_i}{\sum_{i=1}^n \gamma_i}; i = 1, 2, \cdots, n \\ I_i = -\log_2 h_i; i = 1, 2, \cdots, n \end{cases} \quad (4)$$

(3)对上述获取特征信息量的贡献率 ω_i 展开计算,并根据计算结果,构造激光数据特征信息量加权矩阵 W :

$$\begin{cases} \omega_i = I_i / \sum_{i=1}^n I_i \\ W = \text{diag}[\omega_1, \omega_2, \dots, \omega_n] \end{cases} \quad (5)$$

式中, 矩阵对角线元素描述成 diag 形式。

(4) 利用信息量加权矩阵对上述主成分分析的特征线性变化矩阵展开信息融合, 获取融合信息后的特征向量变换矩阵 $Y^* = YW$, 实现多波段激光数据特征值的精准表征。

(5) 完成激光数据特征值 $Y^* = (y_1, y_2, \dots, y_n)$ 获取后, 依据降序排列方法将特征值从大到小顺序排序, 找出最大特征值 $y_{\max}$, 从而获取最大特征值对应的特征向量即最大特征矢量 y_ζ 。

2.2 激光数据时间序列优化

不同波段的数据会因为相同的干扰因素而产生相似的变化。时间序列分析可以帮助识别交叉敏感性, 通过分析不同波段数据随时间的变化关系, 从而在分类方法的中展开融合。因此, 找到最大特征矢量 y_ζ 后, 基于区间数理论, 完成激光数据时间序列的建立。

将获取的最大特征矢量 y_ζ 作为多波段激光数据属性, 并令数据属性与变化状态之间存在正比例函数关系, 计算数据在 t 状态下的绝对变化率 e_{x_t} :

$$e_{x_t} = \left| \frac{(x_t - x_{t-1})(x_{t-1} - x_{t-2})}{x_t} \right| \times y_\zeta \quad (6)$$

式中, t 时段激光数据表述成 x_t 形式。

根据上述获取的数据变化率, 设定数据模糊半径为 L_{x_t} , 获取数据属性 y_ζ 的区间自觉模糊数 B_{x_t} , 完成激光数据时间序列 $F(t)$ 的构建, 过程如下式所示:

$$F(t) = (B_{x_1} \cdot e_{x_1}, B_{x_2} \cdot e_{x_2}, \dots, B_{x_t} \cdot e_{x_t}, \dots, B_{x_T} \cdot e_{x_T}) \quad (7)$$

完成激光数据时间序列构建后, 为进一步增加后续多波段激光数据的分类精度, 使用遗传算法对激光数据时间序列 $F(t)$ 实施优化处理。

使用遗传算法优化数据时间序列时具体流程如下:

步骤 1: 基于激光数据时间序列 $F(t)$, 使用遗传算法展开种群初始化处理, 并使用自然二进制编码模式完成编码。

步骤 2: 设定遗传算法中染色体基因位对应时间序列矩阵列向量, 染色体长度对应矩阵列向量数量。

步骤 3: 在种群大小为 N 时, 采用最小化目标函数构建适应度函数 f , 过程如下式所示:

$$f = \frac{F(t) \cdot e_{x_t} \cdot y_\zeta}{M} \quad (8)$$

式中, M 表示激光数据时间序列总数量。

步骤 4: 基于上述构建的适应度函数计算种群的个体适应度值, 并根据计算结果对种群个体实施选择、交叉、变异操作, 当结果达到设定终止条件时, 算法结束, 输出最优种群 $F'(t)$ (即多波段激光数据最优时间序列)。

3 多波段激光数据分类方法设计

不同波段的激光能够穿透或反射不同深度的物质表面, 从而提供关于目标的不同层次的信息。通过最优时间序列 $F'(t)$ 的获取, 可以更好地识别和分类不同的目标, 以减少其对分类结果的负面影响。模糊分割算法能够处理这些模糊性和不确定性, 在上述获取的多波段激光数据时间序列 $F'(t)$ 中引入模糊逻辑, 对其实施分割处理, 基于分割结果即可实现多波段激光数据的精准分类。

由于提取的多时段激光数据时间序列 $F'(t)$ 为多元时间序列, 表述成 $F'(t) = (f_1, f_2, \dots, f_T)$ 形式, 其中时间序列长度描述成 T 形式, 数据 i 时刻多元变量观测值描述成 f_i 形式。为了在激光数据分类过程中体现时间序列前后数据之间关联性, 以时间序列的时间点为基准, 向前截取时间长度为 ω 段落作为子序列, 记作 $F_i = [f_{i-\omega+1}, \dots, f_{i-1}, f_i]$ 。

设定激光数据的原始时间序列能够被划分成 K 个类别, 令其中 j 类数据段序号集合为 P_j , 将数据的分类定义成一个马尔科夫随机场, 并将随机场中的边看作数据两个变量之间偏相关性。

设定不同类别数据的协方差逆矩阵为 Θ_i , 将其用来表述不同聚类的马尔科夫随机场。矩阵由 $\chi \times \chi$ 个子矩阵构成, 子矩阵尺度设定成 $n \times n$, 其中数据段数量描述成 n , 数据时间窗大小描述成 χ 。因为不同时刻数据之间差别与时间相关, 可将不同时刻逆协方差矩阵互为转置, 因此联合模糊理论构建不同数据类别的逆协方差矩阵 Θ_i , 矩阵表述形式如下式所示:

$$\Theta_i = [A^{(0)}, A^{(1)}, \dots, A^{(\chi-1)T}] \quad (9)$$

式中, 数据模糊隶属度表述成 $A^{(0)}$ (即同一时刻不同数据之间相关关系用子矩阵)。

将激光数据分类问题简化成数据时间序列的分割优化问题, 优化问题即为求解 K 个逆协方差矩阵

$\Theta = \{\Theta_1, \Theta_2, \dots, \Theta_K\}$ 以及最终结果聚类集 $S = \{S_1, S_2, \dots, S_K\}$, 以此构建激光数据时间序列模糊分割的目标函数:

$$\psi = f_i \sum_{i=1}^K [\|\alpha \circ \Theta_i\|_1 - \beta l(X_i, \Theta_i)] \quad (10)$$

式中, 数据点 X_i 的类别对数似然值描述成 $l(X_i, \Theta_i)$ 形式, 元素乘积的范数表述成 $\|\alpha \circ \Theta_i\|_1$ 形式, 正则化参数描述成 α 形式, 一致性系数描述成 β 形式。

通过 K-means 算法, 在时间序列段中选取聚类中心, 计算数据点与聚类中心之间距离, 完成数据分类。过程如下:

(1) 基于上述获取的激光数据时间序列段, 选取 L 个初始质心, 作为时间序列段分类中心。

(2) 对于不同时间序列段, 计算序列点与质心之间欧氏距离 d , 并根据结果将序列段分配给不同的质心, 完成多波段激光数据的有效分类。过程如下式所示:

$$d = \psi \sqrt{(x - x_1)^2 (y - y_1)^2} \quad (11)$$

式中, 不同序列点位置表述成 (x, y) 形式。

4 实验与结果分析

为了验证上述多波段激光数据分类方法的整体有效性, 需要对此方法测试。测试过程中将考虑时间序列模糊分割的多波段激光数据分类(所提方法)、基于随机森林算法的底栖动物高光谱数据分类方法研究(文献[3]方法)、结合物理与几何特性的机载 LiDAR 数据分类方法(文献[5]方法)作为测试方法, 开展对比测试, 验证所提方法的在激光数据分类时的可行性。

设定实验时, 激光数据波段为蓝色、红色、以及近红外、短波红外五种波段, 对应波长分别设定成 450 nm、650 nm、850 nm、1050 nm。利用 Jcp45160z-002 型号的激光扫描仪扫描 450 nm 波段蓝色激光,

激光光束为 13 线 + 13 线 + 7 线, 线条均匀, 清晰明亮, 调制频率为 100 kHz, 占空比为 0 ~ 100 %; 利用光束发散角为 < 1.5 mrad、光束直径为 < 2.3 mm 的孚光精仪扫描 650 nm 激光波长; 利用 850 nm 激光器 1102 型号的波段激光扫描仪扫描 850 nm 波段近红外激光, 采用 OHSP350IR 红外光谱辐照强度分析仪, 用于 1050 nm 波段红外激光检测。

将各类激光扫描仪采集的数据传输至计算机中, 场景图如图 1 所示。



图 1 激光扫描仪数据传输场景图

Fig. 1 Laser scanner data transmission scene diagram

如图 1 所示, 将来自不同型号激光扫描仪(如 Riegl VZ-4000, Trimble TX8, 以及 Velodyne HDL-64E 等)采集的原始激光点云数据, 通过传输速率可达 1Gbps 的 Gigabit Ethernet 高速以太网接口, 以每秒数百万点的速度, 实时传输至高性能计算机中。这些数据包含了不同波段的反射强度信息, 以及每个点的三维坐标 (X, Y, Z) , 确保了数据的完整性和准确性, 为后续的数据处理与分类分析奠定了坚实基础。

4.1 激光数据序列优化结果验证

由于激光数据存在多种不同波段, 受到噪声、遮挡、反射率变化等因素的影响, 都会影响数据序列的质量。验证数据序列构建的效果可以帮助确保数据在分类中没有引入额外的误差。因此, 利用所提方法、文献[3]方法以及文献[5]方法开展多波段激光数据分类时, 验证激光数据序列构建的效果, 测试结果如图 2 所示。

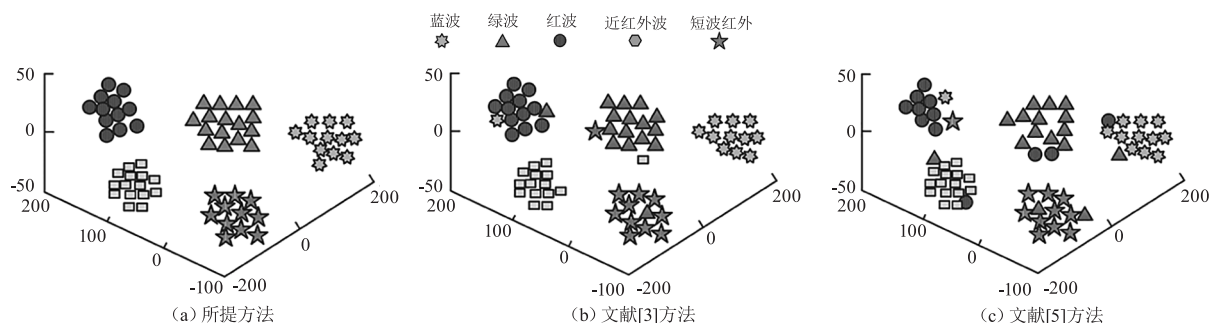


图 2 激光数据序列优化结果对比

Fig. 2 Comparison of the laser data sequence optimization results

分析图 2 可知,所提方法由于在激光数据分类时,通过主成分分析法获取了数据特征最大矢量,通过找到数据中的主要成分(即特征最大矢量),可以有效地减少数据的维度,同时保留大部分原始数据的信息,从而建立了激光数据的时间序列,因此该方法在激光数据分类时,可以提高分类精度。

4.2 抗干扰性能测试

由于多波段激光数据采集时,会不可避免的受到高斯噪声影响,从而令数据中存在噪声数据,影响分类方法的分类性能。为此,利用所提方法、文献[3]方法以及文献[5]方法开展多波段激光数据分类时,利用 LoadRunner 工具对不同方法的分类抗干扰性能展开测试,测试时 LoadRunner 具体运行参数设置如下:

- (1)虚拟用户数:50
- (2)工具运行迭代次数:3
- (3)网络延迟:50 ms;
- (4)带宽限制:1 Mbps;
- (5)最大、最小思考时间:5 s、10 s。

不同方法的分类抗干扰性能测试结果如表 1 所示。

分析表 1 可知,所提方法能够在激光数据分类时,有效识别其中噪声数据并剔除,因此该方法在数据分类时,分类性能好。因为所提方法引入了模糊

分割算法,允许数据点在不同类别之间有一定的隶属度,而不是强制性地将其归类为某一类别。这种模糊性处理有助于识别和处理噪声数据,因为噪声数据通常表现出模糊的特征,难以明确归类。

表 1 不同方法的分类抗干扰性能测试结果

Tab. 1 Classification and anti-interference performance test results of different methods

待分类数据 总数量/个	噪声 数据/个	实际噪声数据去除数量/个		
		文献[3] 方法	文献[5] 方法	所提方法
1000	1	1	1	1
2000	8	8	8	8
3000	15	15	15	15
4000	23	23	23	23
5000	30	30	28	30
6000	41	41	36	41
7000	56	53	50	56
8000	65	60	57	65
9000	79	72	69	79
10000	91	85	82	91

4.3 内存占用测试

开展多波段激光数据分类时,会耗费大量的资源,为此针对上述 3 种方法在激光分类时,对不同方法分类过程中的内存占用量展开测试,以此验证不同方法分类资源消耗量。结果如图 3 所示。

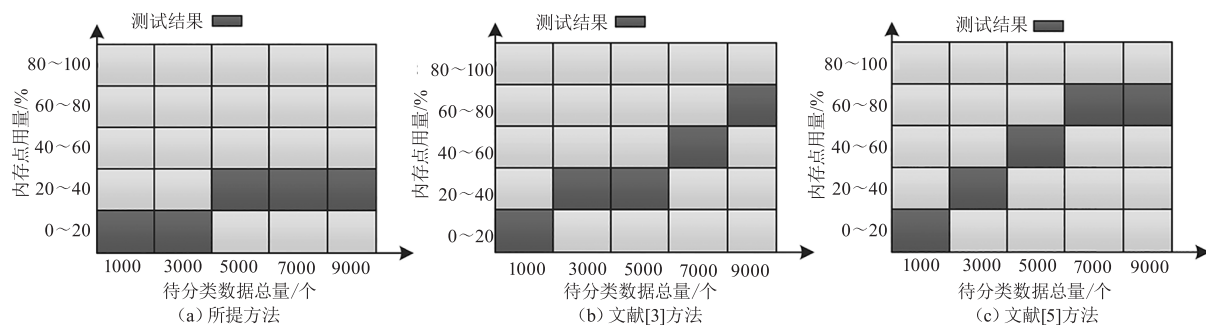


图 3 不同方法的分类过程内存占用量测试结果

Fig. 3 Test results of the classification process memory usage for different methods

分析图 3 可知,所提方法由于在分类时,基于模糊分割理论,对数据时间序列实施了分段处理,缩减了后续分类运行量,因此该方法在后续分类时,有效降低了内存的占用量。

5 结 语

针对传统数据分类方法中存在的问题,提出考虑时间序列模糊分割的多波段激光数据分类。该方法首先提取激光数据特征值,建立时间序列;再联合

模糊分割算法对时间序列分割处理,获取时间序列段;最后通过对时间序列段聚类,实现激光数据的多波段分类。该方法由于在时间序列获取时,过程复杂程度较高,今后有针对该项问题继续优化该方法。

参考文献:

- [1] Cao Shijie, Yao Longying, Zhang Zhiyi. Scanning inner surface of object-based on ring laser[J]. Computer Simu-

- lation,2023,40(6):295-300. (in Chinese)
- 曹世界,姚隆兴,张志毅. 基于环形激光的物体内表面扫描[J]. 计算机仿真,2023,40(6):295-300.
- [2] Xū Qiang,Guo Chen,Dong Xiujun. Application status and prospect of aerial remote sensing technology for geohazards[J]. Acta Geodaetica et Cartographica Sinica, 2022,51(10):2020-2033. (in Chinese)
- 许强,郭晨,董秀军. 地质灾害航空遥感技术应用现状及展望[J]. 测绘学报,2022,51(10):2020-2033.
- [3] Dong Jianjiang,Tian Ye,Zhang Jianxing, et al. Research on the classification method of benthic fauna based on hyperspectral data and random forest algorithm[J]. Spectroscopy and Spectral Analysis, 2023, 43(10):3015-3022. (in Chinese)
- 董建江,田野,张建兴,等. 基于随机森林算法的底栖动物高光谱数据分类方法研究[J]. 光谱学与光谱分析,2023,43(10):3015-3022.
- [4] Dubey S,Kumar V P,Prasanna K S. Logistic regression analysis for the classification of human pulse data[J]. ECS transactions,2022,107(1):19739-19748.
- [5] Zhao Yiqiang,Zhang Qi,Liu Changlong, et al. Airborne LiDAR data classification method combining physical and geometric characteristics[J]. Infrared and Laser Engineering,2023,52(11):1-13. (in Chinese)
- 赵毅强,张琦,刘长龙,等. 结合物理与几何特性的机载 LiDAR 数据分类方法[J]. 红外与激光工程,2023,52(11):1-13.
- [6] Xue Zhanao, Li Yongxiang, Yao Shouqian, et al. Data classification method based on Bayesian intuitionistic fuzzy rough sets[J]. Journal of Shandong University:Natural Science,2022,57(5):1-10. (in Chinese)
- 薛占熬,李永祥,姚守倩,等. 基于 Bayesian 直觉模糊粗糙集的数据分类方法[J]. 山东大学学报:理学版,2022,57(5):1-10.
- [7] Wang Rui,He Zhilong,Zhang Zhen, et al. Evaluation of the effect of nitrogen forms on the growth of camellia oleifera seedling based on principal components analysis and membership function method[J]. Journal of Central South University of Forestry & Technology,2023,43(10):11-19. (in Chinese)
- 王瑞,何之龙,张震,等. 基于主成分分析及隶属函数法评价氮素形态对油茶苗木的影响[J]. 中南林业科技大学学报,2023,43(10):11-19.
- [8] Zhang Xiaodi,Liu Jiyun,Wei Mingyue, et al. Method for calculating the state image pose of the power panel cabinet pressing plate based on multi-band laser data clustering[J]. Laser & Infrared,2023,53(7):1110-1116. (in Chinese)
- 张笑迪,刘吉昀,韦鸣月,等. 基于多波段激光数据聚类的电力屏柜压板状态图像位姿解算方法[J]. 激光与红外,2023,53(7):1110-1116.