

高斯混合聚类对移动曲面拟合滤波分类的应用

邢承滨^{1,2}, 龚声胜^{1,2*}, 于晓亮^{1,2}, 李易馨³

1. 天津水运工程勘察设计院有限公司, 天津 300456;
2. 交通运输部天津水运工程科学研究所, 天津 300456;
3. 长沙理工大学 经济与管理学院, 湖南 长沙 410014)

摘要: 为了提高激光雷达点云滤波算法的精度和自适应性,对移动曲面滤波算法进行改进。采用格网边界点构建曲面约束条件,检验格网内是否全部为建筑物点。利用区域拟合求解地形的起伏,引入机器学习中高斯混合模型(GMM)对地形起伏进行滤波分类,将移动曲面中的种子点作为聚类算法中的靶向点参与分类学习。实验数据为雷达飞行的自测区,对于自测区采用随机抽样的方式,检验判断滤波效果。同时为检验 GMM 算法的准确性,在三类误差检验方式的基础上,增加了 Kappa 系数作为检验方式。通过与谱系聚类分类算法对比,证明所提算法能取得较好的滤波效果。

关键词: 滤波分类; 激光雷达; 高斯混合聚类模型; 最大期望; 降维

中图分类号: P237 **文献标志码:** A **DOI:** 10.3788/IRLA20200501

Application of Gaussian Mixture Clustering to moving surface fitting filter classification

Xing Chengbin^{1,2}, Gong Shengsheng^{1,2*}, Yu Xiaoliang^{1,2}, Li Yixin³

1. Tianjin Survey and Design Institute for Water Transport Engineering Co., Ltd., Tianjin 300456, China;
2. Tianjin Research Institute for Water Transport Engineering, Ministry of Transport of the People's Republic of China, Tianjin 300456, China;
3. School of Economics and Management, Changsha University of Science and Technology, Changsha 410014, China)

Abstract: In order to improve the accuracy and adaptability of the LiDAR point cloud filtering algorithm, an improved moving surface filtering algorithm had proposed. The boundary points of the grid were used to construct the surface constraint conditions to test whether all the building points in the grid. The area fitting was used to solve the terrain fluctuations. The Gaussian Mixture Model (GMM) in machine learning was introduced to filter and classify the terrain undulations, and the seed points in the moving surface were used as the target points in the clustering algorithm to participate in the classification learning. The experimental data was the self-test area of radar flight. The filtering effect of the self-test area was tested and judged with random sampling. At the same time, the Kappa coefficient was added as the test method to test the accuracy of the GMM algorithm on the basis of the three types of error test methods. Compared with the pedigree clustering classification algorithm, it is proved that the proposed algorithm can achieve better filtering effect.

Key words: filtering classification; LiDAR; Gaussian Mixture Clustering model; maximum expectation; dimensionality reduction

收稿日期:2020-12-18; 修订日期:2021-02-24

基金项目:中央级公益性科研院所基本科研业务费专项资金(TKS20200318, SJY20200101)

作者简介:邢承滨,男,硕士生,主要从事机载雷达点云算法方面的研究。

通讯作者:龚声胜,男,高级工程师,硕士,主要从事海洋测绘技术方面的研究。

0 引言

激光雷达 (Light Detection and Ranging, LiDAR) 是当前流行的一种主动获取点云数据的手段。经过 20 多年的发展,激光雷达的硬件技术已经取得长足的进步,当前流行的机载平台有 Leica³(ALS80-UP), Optech⁴(Titan),北京天绘⁵(AP-3500) 等,都采用 1 064 nm 的波长。激光雷达主要面向中小型区域地形植被、建筑物、水下地形等三维数据获取^[1-3]。

激光雷达系统测量原理是利用发射激光束测量与目标之间距离,结合发射器位置和姿态联合解算,得到目标区域的三维点云 (X,Y,Z)。相较于传统航空摄影,激光雷达不仅能够全天候作业,并且抗干扰能力较强,点云数据量丰富。当前点云数据是继三大传统地理空间产业 (GNSS, GIS, RS) 后的第四大产业。但是点云数据也存在缺陷,例如数据量庞大,高冗余,采样粒度分布不均等^[4-7]。

点云数据的后处理一直是难题,其中雷达点云滤波 (雷达点云地面点和非地面点的分离称为滤波) 更是数据处理的关键。点云滤波技术发展已有几十年的历史,众多学者提出了多种滤波算法,并且不断完善。如今滤波算法主要分为四类: (1) 形态学滤波算法^[8-11]; (2) 基于坡度的滤波算法^[12]; (3) 基于曲面拟合的滤波算法^[13-14]; (4) 基于聚类分割的滤波算法^[15]。

移动曲面滤波算法最早由武汉大学的张小红提出,文中基于传统的移动曲面算法,引入机器学习中高斯混合模型 (Gaussian Mixture Model, GMM) 聚类算法对地形进行分类,为检验文中算法的分类效果,与谱系聚类算法 (另一种常见的机器学习聚类方法) 进行实验对比,判断两种机器学习算法的滤波效果。

1 约束条件下移动曲面滤波算法

移动曲面算法的原理将测区划分为规则格网,采用移动窗口的方式对所有数据点进行遍历。对规则格网进行种子点选择,利用种子点拟合地面模型,求解真实数据点的高程与拟合高程之间的高差。

1.1 点云数据预处理

文中采用直方图方法对数据进行粗差剔除,初步剔除粗差有利于种子点的准确选择。对于测区进行格网划分,如图 1(a) 所示, i, j 表示测区格网建立的行

列号。横纵步长相等,即 $dx=dy$ 。建立格网的目的是保证每一个数据点都有唯一索引号,同一格网内数据点索引号一致的。

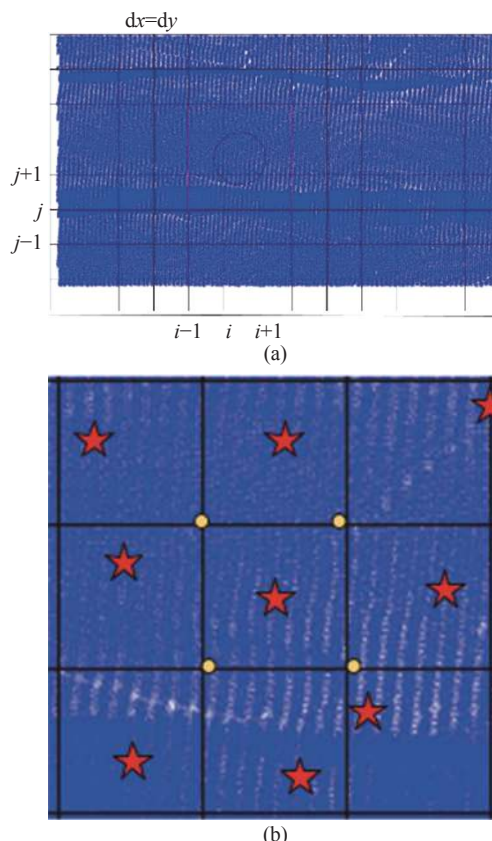


图 1 测区格网划分和种子点选取

Fig.1 Grid division of survey area and selection of seed points

文中采用大小 3×3 的移动格网,单个格网内最低点作为种子点 (再经过剔除粗差的数据区域,格网内最低点通常确定为地面点)。移动窗口能较好覆盖建筑区域,并且格网尺寸不宜太大,尽量简化格网内地物种类,为后续建立分类做准备。如图 1(b) 所示,星号点为移动格网内种子的地面点,利用种子点建立初始地面。种子点的选择对于文中算法至关重要,不仅初期进行曲面拟合,后期更会作为靶向点参与地面点的分类。

1.2 初始地面模型建立

文中采用二次曲面拟合方式建立初始地面模型。

$$z = ax^2 + by^2 + cxy + dx + ey + f \quad (1)$$

式中: z 为种子点的高程; x, y 为种子点的平面坐标; $a \sim f$ 为拟合方程的参数。利用观测值建立误差方程:

$$\begin{cases} z_1 = a_0 + a_1x_1 + a_2y_1 + a_3x_1^2 + a_4x_1y_1 + a_5y_1^2 \\ z_2 = a_0 + a_1x_2 + a_2y_2 + a_3x_2^2 + a_4x_2y_2 + a_5y_2^2 \\ \vdots \\ z_n = a_0 + a_1x_n + a_2y_n + a_3x_n^2 + a_4x_ny_n + a_5y_n^2 \end{cases} \quad (2)$$

利用最小二乘原理, 求解上述二次曲面的 6 个参数, 再利用求解的拟合参数建立地面模型。如图 2(a) 所示, 为某格网内数据真实高程与拟合高程的分布。图 2(b) 表示真实高差与拟合高差的分布。通常, 地面点拟合高程与真实高程近似, 地物点拟合高程与真实高程差异较大, 高差为地物实际高度, 这是为后续分类提供重要理论。

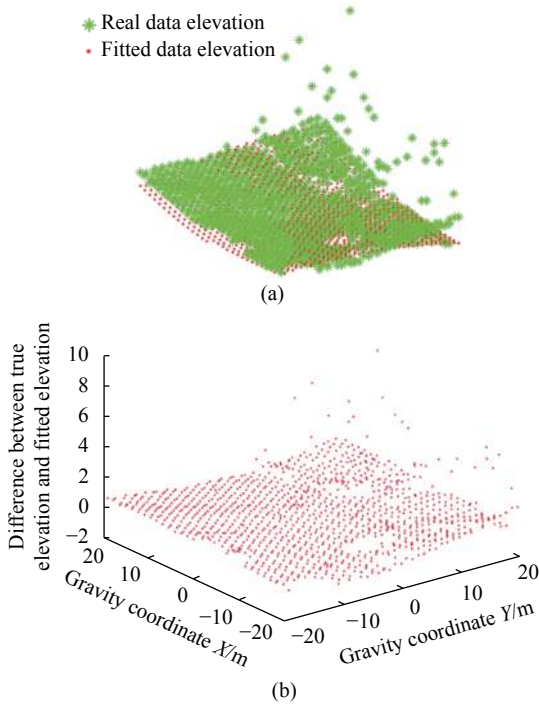


图 2 点云真实高程与拟合高程分布

Fig.2 Point cloud real elevation and fitting elevation distribution

1.3 约束条件判断种子点的准确性

多种滤波算法将格网内最低点定义为种子点, 但是存在部分特殊情况, 例如格网内地形都为建筑的特殊地形, 后续的分类将建筑点全部误分类为地面点, 造成对数据的错误判断。对于这种特殊地形, 最低点判断为种子点就会存在误差。建筑地区有着明显的特征, 边界存在明显高差, 建筑区域内高程起伏较小, 这也是与地面点容易混淆的特征。文中基于建筑地区的特征, 提出角度阈值检验算法判断格网内是否全

为建筑点。

如图 3 所示, a, b, c, d 为同一格网中的边界点, 判断最低点与模拟平面的角度, 经过众多实验判断出 $\max\{\theta_1, \theta_2, \theta_3, \theta_4\} \leq 5^\circ$, 表示平面地形较为平滑。对于较为平坦的地形, 进行二次判断。二次判断采用梯度差方式进行。建筑顶端表面平整, 将格网内平均高程与周围临近 4 个格网内平均高程逐次比较。经过大量实验计算, 平均高程与周围邻近格网平均高程阈值为 8 cm, 对建筑地区滤波效果最佳, 阈值范围内该格网定义为建筑格网, 格网内全部定义为地面点。当前还没获取一种高效的科学计算手段判断角度阈值与梯度阈值, 判断方式还停留在统计计算。阈值设定主要判断测区内是否完全为建筑点, 这种验证方式只对存在大量建筑的区域效果明显, 是一种先验条件下的判别方式; 对于非密集建筑区意义不大。通常所选测区都比较复杂, 先验一下测区数据, 避免格网全部为建筑点造成的误差。

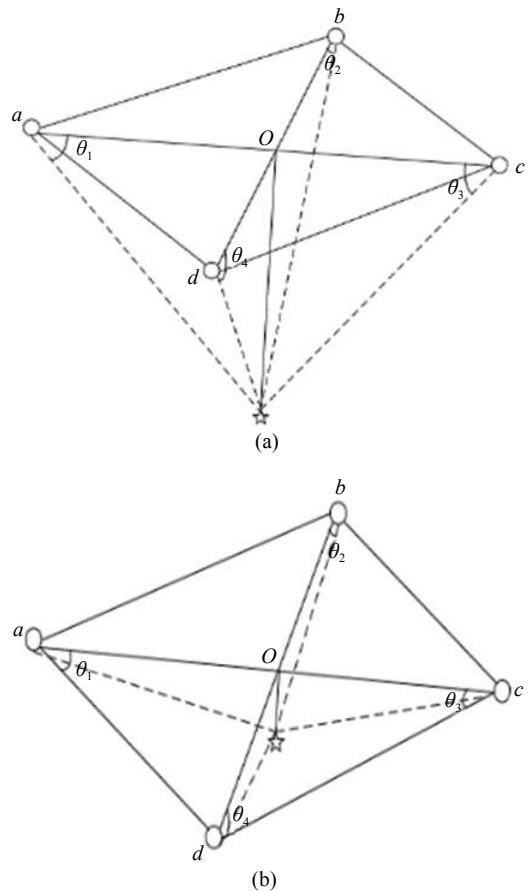


图 3 角度阈值法检验建筑点

Fig.3 Angle threshold method to inspect building points

2 GMM 聚类算法对于点云的聚类分析

上文提出地面点拟合高程与真实高程近似, 地物点拟合高程与真实高程之间高差近似为地物高度。如图 4 所示, A_1 为地面点, B_2 为相邻地物点 (非地面点), A_1B_2 水平距离为 Δd 。 Δh_2 为 A 点与拟合值高差, Δh_1 为 B 点与拟合值高差。 AB 两点的地形起伏为 $\Delta h_1 + \Delta h_2$ 。 相邻点的水平距离和高差距离作为 GMM 聚类算法的数据基础。将三维数据降维为二维起伏数据, 利用 GMM 聚类算法进行学习判断。

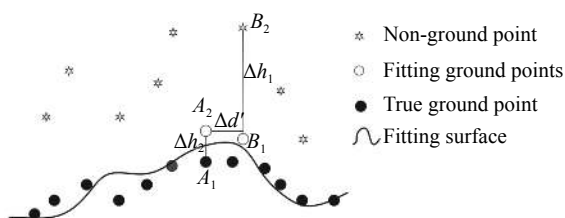


图 4 三维数据的降维处理

Fig.4 Dimensionality reduction processing of 3D data

2.1 GMM 聚类算法

GMM 算法是一种在机器学习中广泛使用的聚类方式。GMM 算法原理是将事物分解为若干的基于高斯概率密度的函数, 每一个混合成分对应一个高斯分布, 任意形状的概率分布都可以用多个高斯分布函数近似表达^[16-17]。GMM 算法原理概率模型为:

$$P(y|\theta) = \sum_{k=1}^K a_k \phi(y|\theta_k) \quad (3)$$

式中: K 为最大高斯概率密度函数的个数; a_k 为每一个高斯分布的权重, $a_k \geq 0$; $\phi(y|\theta_k)$ 为第 k 个高斯分布的概率密度函数; y 为观测值。其中 $\theta = (a_1, a_2, \dots, a_k; \theta_1, \theta_2, \dots, \theta_k)$, $\theta_k = (\mu_k, a_k^2)$, $k = 1, 2, \dots, k$, θ 为未知参数, 包括权重 a_k 与 θ_k , θ_k 包括期望 μ_k 和方差 a_k^2 , 高斯混合 k 个模型需要估计 3 k 个参数。

$$\phi(y|\theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(y-\mu_k)^2}{2\sigma_k^2}\right) \quad (4)$$

公式 (4) 中, 每个观测样本出现的概率表示为 k 个高斯分布的加权。聚类的核心思想是对于某个样本 y_j , 将该样本代入到 k 个高斯分布中求出属于每个类别的概率, 公式 (5) 表示第 k 个类别概率:

$$\frac{a_k \phi(y_j|\theta_k)}{\sum_{k=1} a_k \phi(y_j|\theta_k)} \quad (5)$$

2.2 GMM 模型中关键参数的 EM 算法

EM(Expectation-Maximization) 算法又称最大期望算法, 要估计的参数 $\theta = (a_1, a_2, \dots, a_k; \theta_1, \theta_2, \dots, \theta_k)$, $\theta_k = (\mu_k, a_k^2)$, $k = 1, 2, \dots, k$ 。采用极大似然估计参数 θ , 使得观测值数据 y 的对数似然函数 $L(\theta) = \log p(y|\theta)$ 极大化^[18-19]。公式 (6)、(7) 为建立的似然方程:

$$\theta^* = \arg \max_{\theta} L(\theta) \quad (6)$$

$$L(\theta) = \log P(y|\theta) = \sum_{j=1}^N \left[\log \left(\sum_{k=1}^k a_k \phi(y_j|\theta_k) \right) \right] \quad (7)$$

根据 EM 算法求解参数 θ 。可以获得观测值 y , 但是确定观测数据来自第 k 个模型的数据却是未知, 用隐函数变量 z 表示。EM 算法的迭代过程分为 E 步和 M 步:

- (1) 选择参数的初始值 θ_0 开始进行迭代;
- (2) E(Expectation) 步: 记第 n 次迭代参数为 θ_n , 因此 $n+1$ 的 E 步为:

$$Q(\theta|\theta_n) = \sum_{i=1}^n \sum_z p\theta_n(z|y_i) \log p^{\theta}(y_i, z) \quad (8)$$

Q 函数是指在给定观测数据 y 和第 n 次迭代的参数 θ_n 时, 对数似然函数的 $\log p^{\theta}(y, z)$ 的期望;

- (3) M(Maximization) 步: 求得 $Q(\theta|\theta_n)$ 最大化的 $\theta(n+1)$, 即确定第 $n+1$ 次的模型参数;

$$\theta_{n+1} = \arg \max_{\theta} Q(\theta|\theta_n) \quad (9)$$

- (4) 重复步骤 (2)、(3) 直到收敛;

利用 EM-高斯混合模型聚类求解出参数 θ , θ 参数中包含均值向量 μ_i 和协方差矩阵。

计算期望值 μ_i' :

$$\mu_i' = \frac{\sum_{k=1}^m p(z_k = i|y_k) \cdot y_k}{\sum_{k=1}^m p(z_k = i|y_k)} \quad (10)$$

计算协方差矩阵 σ_k^2 为:

$$\sigma_k^{2'} = \frac{\sum_{k=1}^m p(z_k = i|y_k)(y_k - u_i')^T}{\sum_{k=1}^m p(z_k = i|y_k)} \quad (11)$$

计算混合系数 a_k' 为:

$$a_k' = \frac{\sum_{k=1}^m p(z_k = i|y_k)}{m} \quad (12)$$

式中: Z_k 为第 k 个高斯模型。利用最大期望迭代求解出高斯混合模型的权重和期望方差,再利用高斯混合

模型判断点的聚类结果。

2.3 种子点作为靶向点进行滤波分类

高斯聚类结果不能获取数据的属性,即聚类只是表示同一类型点的集合,并不能显示地面点为哪一类。因此对于地面点类型需要新的判断方式,上文提及格网内种子点为地面点,种子点所在类型就是地面点。种子点作为靶向点参与滤波分类,种子点所在聚类结果中的类型判定为地面点类型。EM-高斯混合模型对于移动曲面算法改进算法流程如图 5 所示。

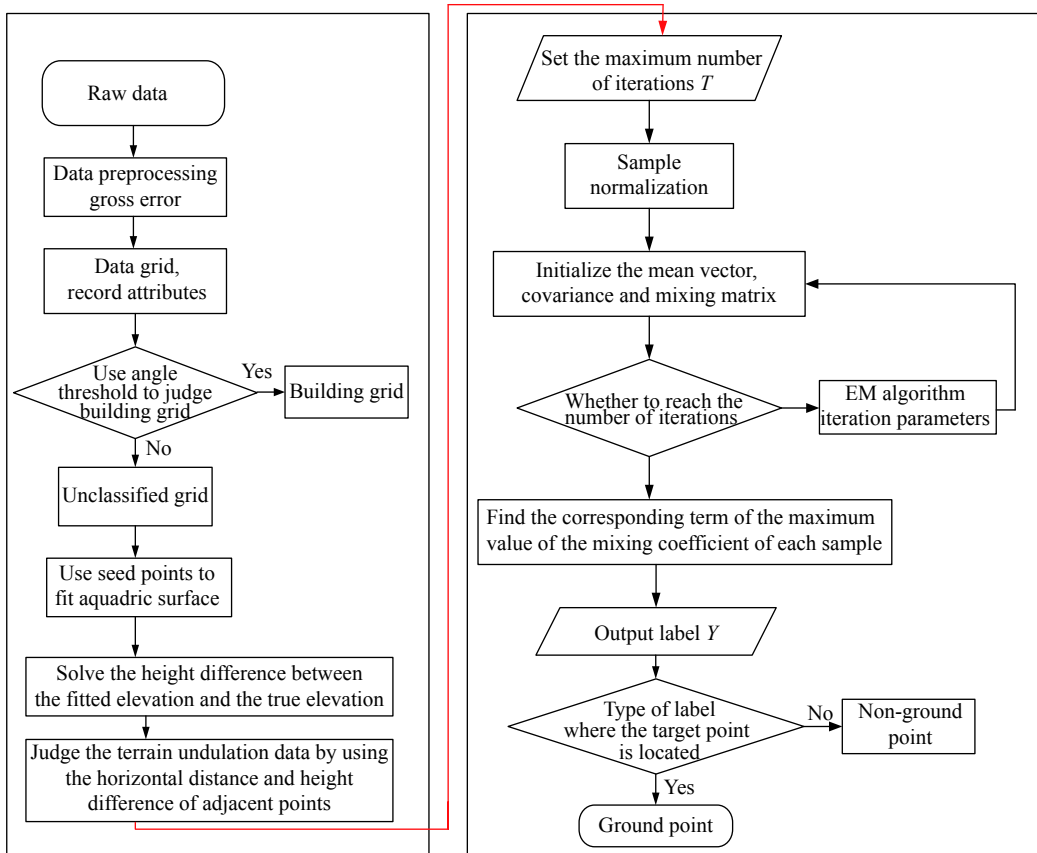


图 5 EM-GMM 模型的移动曲面算法

Fig.5 Moving surface algorithm of EM-GMM model

3 实验分析

文中采用定性和定量方式检验滤波分类效果,定性分析中采取随机抽样的方式检验滤波效果。

3.1 定性实验分析

定性分析主要通过目视方式对滤波效果进行判断,对比滤波前后的地形检验滤波效果。文中实验采用定性分析的数据由华测 AS300 便携机载激光雷达

扫描系统获取。测域为天津某一工业区域,内部设有建筑厂房、主办公楼、湖泊、树木、汽车等小型地物。采用随机抽样的方式判断滤波效果。测区内数据点约为两千万,随机抽取某一格网数据进行展示,实验抽取第 28 格网为例。因文中测区数据平面坐标经过重心化处理, X, Y, Z 会有部分呈现负值。

图 6(a)、(b) 表示测区滤波分类前后点云分布,图 6(c) 表示聚类结果。经过多次实验,通常将高斯混合模型

定义为 4 个。树木、建筑地形、地面点、独立小地物，比如汽车等，此网格为一个包含部分建筑区域的网格。树木呈现为渐变式高差，高差逐渐起伏；建筑物呈现两极分化，分为距离陡变和逐渐变化；地面点由于点云数据量十分庞大，地形起伏较为平缓，绝大多数数据量集中在 0 附近，由于每个格网内存在靶向点，靶向点的数据类型可以直接确定地面点的类型；独立地物变化区域较小。如图 6(a)、(b) 所示，平均高程下降七八米左右，正是测区内建筑物的高度，区域内较高建筑地区点云被剔除。图 6(d)、(e) 表示点云滤波前后的地面模型，对于图 6(d) 中标注的建筑地物很好过滤，地面模型保留较好。图 6(f)、(g) 为等高线表示，经过滤波处理，图 6(g) 中不存在图 6(f) 中明显的建筑物闭合，边界密集的等高线分布。GMM 算法

对于建筑物有着较好的过滤效果。

抽取第 56 个格网进行滤波效果分析。由等高线结合点云分布，测区为一个存在独立地物的斜坡地形。图 7(a) 为聚类后 4 类地物，图 7(b) 为滤波前的地面模型，图 7(c) 为滤波后的地面模型。图 7(b) 中地形可以判断存在独立地物，比如旗杆、汽车等。这类地物变化区域较小，独立地物大小相对有限，文中算法可以很好地过滤其中的独立地物，将图 7(b) 中两处独立地物很好地剔除。根据图 7(d) 中等值线的分布可以判断独立地物的存在。两处相对独立的等高线，而图 7(e) 中不存在这两处区域等高线，只显示逐渐起伏的地形。

虽然随机抽样对于滤波效果判断存在片面性，但是随机抽样方式是对点云类型检验滤波效果的有效

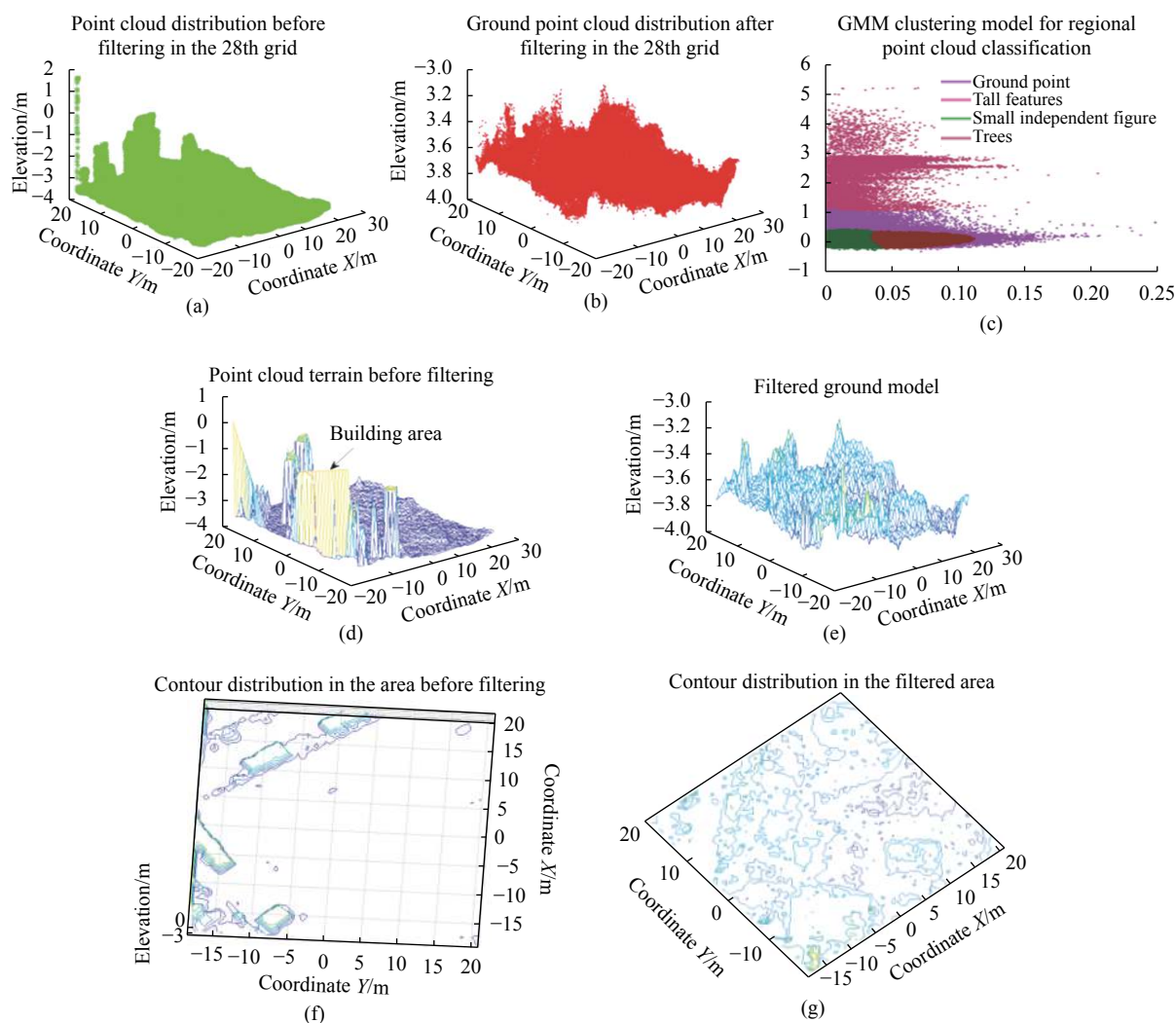


图 6 工业区第 28 格网数据展示

Fig.6 Data display of the 28th grid in the industrial zone

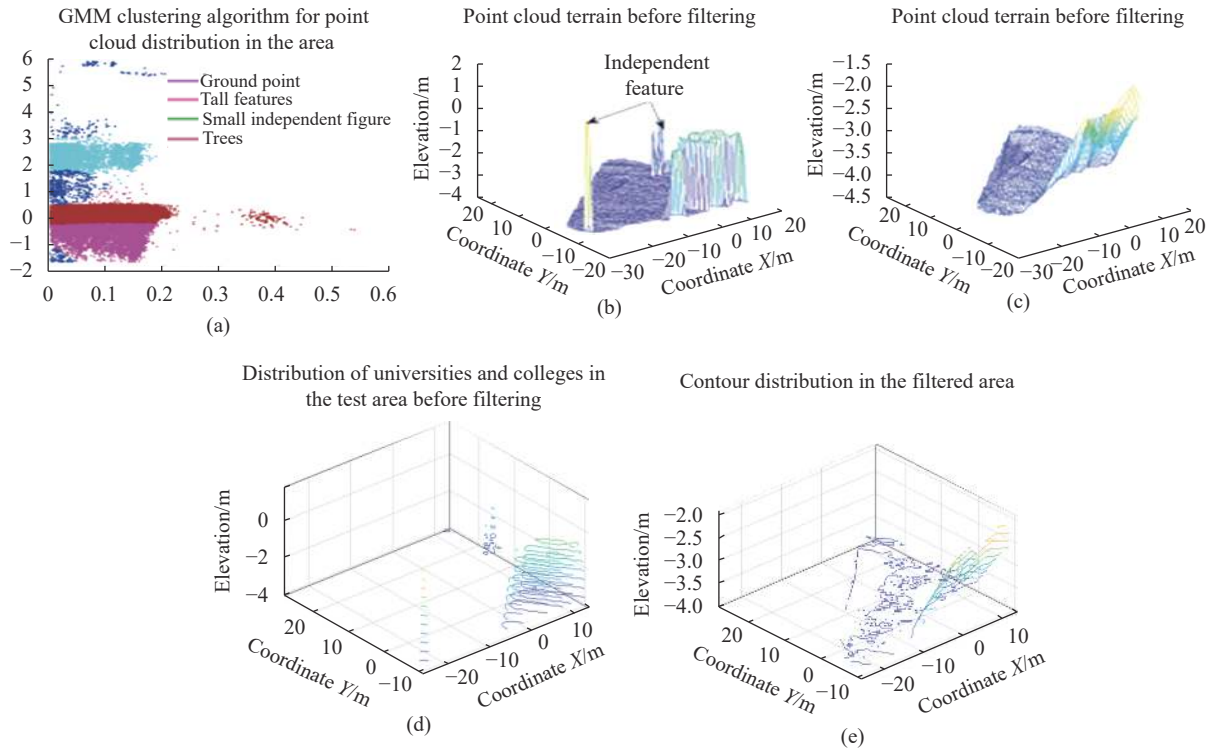


图 7 工业区第 56 格网数据展示

Fig.7 Data display of the 56th grid in the industrial zone

方式, 抽样数据能一定程度反映全部滤波效果。

3.2 定量实验分析

此次定量实验数据采用 2003 年摄影测量与遥感协会 (ISPRS) 公布的雷达点云数据。该数据为 Optech ALTM 机载激光扫描仪获取的 Vaihingen/Enz 和 Stutt-

gart 市的点云数据。其中城市数据: Site1~Site4; 村庄数据: Site5~Site8, 表 1 对每个样本的属性进行详细解释。

滤波误差类型分为 3 种: I 类误差 (Type I error), 将地面点误分为非地面点; II 类误差 (Type II error),

表 1 3 类误差与 Kappa 系数判定方法

Tab.1 Three types of errors and Kappa coefficient judgment method

Test sample	Sample name	Topographic features
Site1	Samp11	Vegetation and buildings on steep slopes
	Samp12	Small features
	Samp21	Narrow bridge
Site2	Samp22	Bridge/passage
	Samp23	Complex and huge buildings, intermittent terrain
	Samp24	Slope
Site3	Samp31	Cluster low-value points (multi-path effect)
Site4	Samp41	Discontinuous terrain
	Samp42	Buildings, high-frequency undulating features
	Samp51	Vegetation on the slope
Site5	Samp52	Low vegetation, fractured steep ridges
	Samp53	Intermittent terrain
	Samp54	Low-resolution building points
Site6	Samp61	Low-resolution building points
Site7	Samp71	Bridge, discontinuous terrain

将非地面点误分为地面点;总误差 (Total error), 所有误分点数/数据点总数^[20]。为增加数据的稳健性测度,文中同时引用了 Kappa 系数。3 类误差与 Kappa 系数的计算公式如表 2 所示。

此次定量实验采用对比分析的方式,引入另一种机器学习算法:谱系聚类算法。两种聚类算法对同一样本进行分类分析,判断哪种算法滤波分类效果更好。如表 3 所示。

表 2 测区地形特征与误差分布

Tab.2 Topographic features and error distribution of the survey area

Total number	Type I error	Type II error	Total error	P_A	P_C	Kappa
$p_e=p_a+p_b+p_c+p_d$	$p_b/(p_a+p_b)$	$p_c/(p_c+p_d)$	$(p_b+p_c)/p_e$	$(p_a+p_d)/p_e$	$[(p_a+p_b)(p_a+p_c)+(p_c+p_d)(p_b+p_d)]/p_e^2$	$(P_A-P_C)/(1-P_C)$

表 3 两类聚类算法对于数据样本误差统计

Tab.3 Two types of clustering algorithms for data sample error statistics

Algorithm type	Gaussian clustering model classification algorithm				Pedigree clustering model classification algorithm			
	Sample name	Type I error	Type II error	Total error	Kappa coefficient	Type I error	Type II error	Total error
Samp11	3.43%	14.38%	5.40%	81.79%	22.25%	6.38%	13.82%	72.00%
Samp12	3.68%	6.25%	5.15%	89.56%	10.06%	6.01%	8.36%	83.01%
Samp21	1.59%	4.52%	2.54%	94.19%	8.89%	8.64%	8.81%	80.37%
Samp22	7.31%	6.15%	6.48%	84.16%	10.81%	22.41%	13.09%	61.71%
Samp23	9.16%	9.81%	9.56%	80.13%	15.38%	16.82%	16.26%	66.47%
Samp24	3.70%	14.2%	7.32%	83.45%	28.82%	10.15%	20.12%	60.15%
Samp31	4.04%	6.24%	5.22%	89.51%	9.18%	12.50%	10.00%	74.48%
Samp41	1.95%	2.17%	2.04%	95.81%	5.61%	3.53%	4.77%	90.18%
Samp42	5.56%	3.87%	4.21%	87.53%	7.42%	7.46%	7.45%	84.46%
Samp51	4.91%	4.84%	4.90%	86.23%	16.39%	8.51%	12.96%	74.01%
Samp52	9.36%	10.53%	9.72%	77.84%	14.67%	7.27%	11.54%	76.76%
Samp53	5.75%	15.79%	6.48%	60.69%	10.27%	34.04%	14.39%	52.59%
Samp54	5.34%	1.30%	3.16%	93.62%	5.13%	12.23%	8.98%	82.03%
Samp61	2.67%	5.88%	2.85%	76.53%	10.94%	0.00%	9.15%	72.69%
Samp71	3.23%	7.25%	4.26%	88.60%	5.29%	12.90%	7.80%	82.26%

定量实验可知,在 15 组样本中,高斯聚类 I, II 类误差均在 10%,表明高斯聚类分类准确度较高,相较于谱系聚类中,部分样本中 I, II 类误差超过 10%,在 Samp24 中,谱系聚类 I 类误差 28.82%,相较于高斯聚类的 3.7%,存在较大误差。由图 8 直方图可以更加直观判断,两类聚类算法的误差分布,另外通过 Kappa 系数对比,高斯聚类稳健性较高。

从 15 组样本中不同样本数据可知,文中聚类算法对于平坦地区分类效果较好,稳健性较高。对于个别间断样本,例如样本 Samp53,误差分布较大并且稳健性较低,高斯聚类在 60.69%,谱系聚类低至 52.59%。

此次算法重点分析两算法统一存在误差最大和

稳健性最差的样本 Samp53。对于 Samp53 样本,采用高斯聚类与谱系聚类进行对比分析。图 9(a)、(d) 分别为高斯聚类和谱系聚类结果显示;图 9(b)、(e) 为两类算法误差分布与地面模型;图 9(c)、(f) 为两类算法分类后的地面模型。由上述可知,Samp53 为间断地形,如图 9(b) 所示, I, II 类误差主要集中在地物点与地面点的边界处。I 类误差过大的主要原因是陡变的地面点,拟合高程与真实高程高差过大,相邻点起伏过大被分为地物点,而边界点作为地面点,造成错误的分类。真实地形与拟合地形在间断地区存在较大差异。

尽管两类滤波对于间断地形都略有不足,但是

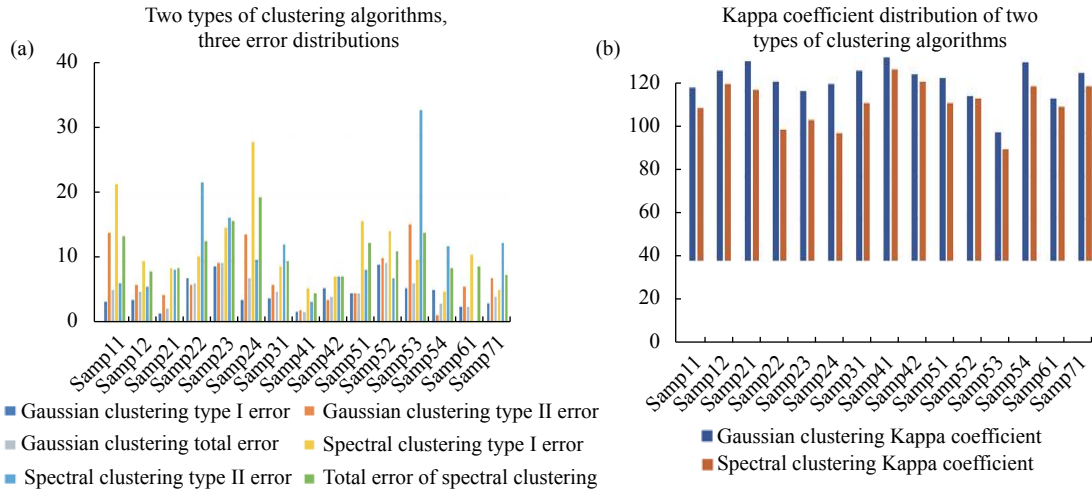


图 8 两类聚类算法三类误差和 Kappa 系数分布

Fig.8 Three types of errors and Kappa coefficient distributions of the two types of clustering algorithms

高斯混合模型与谱系聚类算法针对间断地形分类效果存在较大差异,以图 9(b)中 II 类格网数据为例。图 10(a)~(c)为高斯聚类分类:图 10(a)为根据格网内靶向点判断地面类型,其他属性合并为地物点;图 10(b)为分类后数据属性,图 10(c)为高斯聚类滤波效果。同理,图 10(d)~(f)为谱系聚类分类。

通过对图 10(c)与图 10(f)进行对比发现,高斯聚类中误差数目,尤其是 II 类误差明显小于谱系聚类。

II 类误差主要分布在山脊线, I 类误差主要分布在山体边缘线。

高斯聚类和谱系聚类同样是采用相邻点地形起伏作为聚类依据,造成分类差异的主要原因为高斯聚类里利用高斯曲线模拟地物类型,利用不同类别高斯曲线差异完成分类。高斯聚类的优势在于大量数据分类噪声不敏感,对于距离和密度分类,适合高维特征,同时高斯聚类可以完成较多形状的聚类。

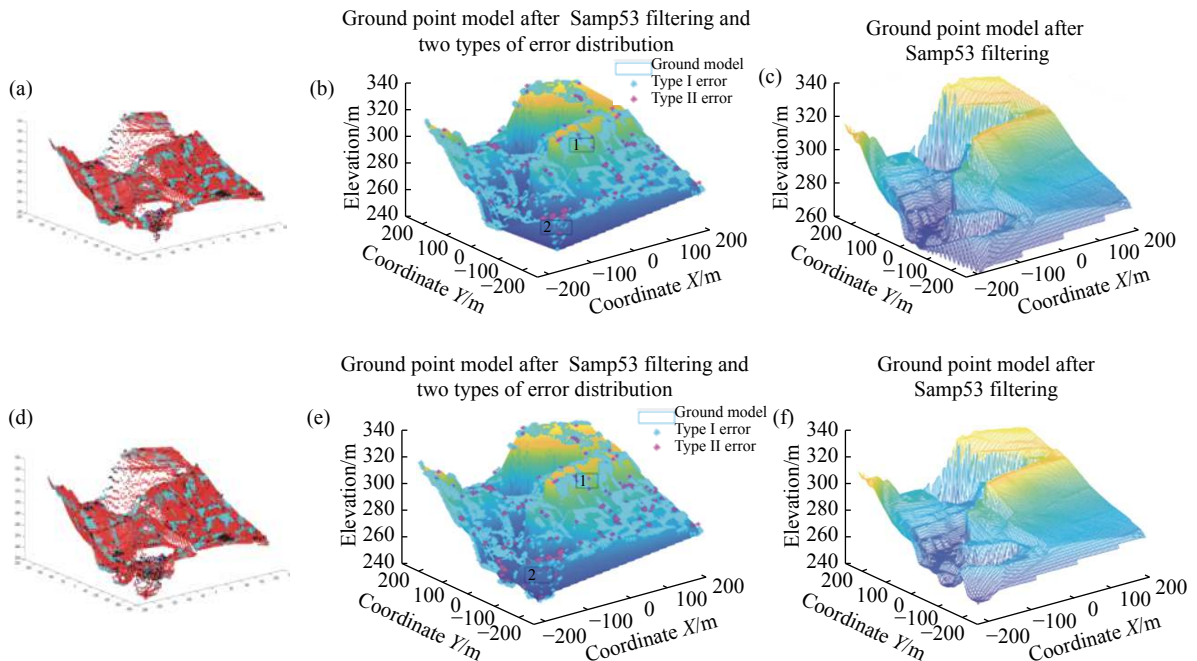


图 9 两类聚类算法对于样本 Samp53 数据处理展示

Fig.9 Two types of clustering algorithms show the data processing of Samp53

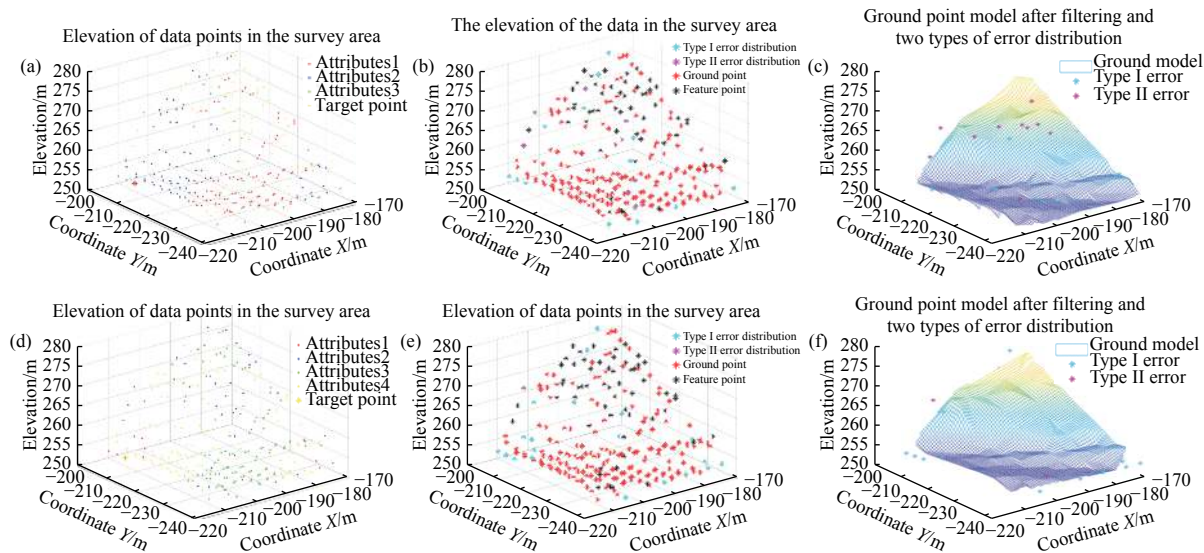


图 10 两类聚类算法对于样本 Samp53 中指定格网数据处理展示

Fig.10 Two types of clustering algorithms show the processing of the specified grid data in Samp53

而谱系聚类对于噪声非常敏感,并且需要在测试前准备判断类别的个数。陡峭地区存在个别噪声,对于谱系聚类算法分类效果影响较大,同时在山体地区,主要地物不再是建筑,而是树木、斜坡等,地物种类不固定于 4 类,谱系聚类按照同一 4 类特征处理,也会造成部分误差。

4 结 论

文中通过移动曲面方式,利用相邻数据点高程起伏作为高斯混合模型分类的依据,对于测区内数据进行分类,利用种子点作为靶向点进行属性判断。通过定性和定量实验分析,该算法可以整体取得较好的滤波效果。传统滤波算法与新型机器学习技术的结合为点云滤波技术的发展提供了新型思路。引入高斯混合模型可以充分利用数据点之间的相关性,对于点云数据拓扑关系不明确的问题有较好改善。对于其中仍存在的问题,例如定义格网数据类型固定为 4 种,部分特殊地形不能用高斯曲线很好拟合等,后续工作将会逐步完善。

参考文献:

[1] Yang Bisheng, Liang Fuxun, Huang Ronggang. Progress, challenges and perspectives of 3D LiDAR point cloud processing [J]. *Acta Geodetica et Cartographica Sinica*, 2017, 46(10): 1509-1516. (in Chinese)

[2] Yan W Y, Shaker A, El-Ashrawy N. Urban land cover classification using airborne LiDAR data: a review [J]. *Remote Sensing of Environment*, 2015, 158: 295-310.

[3] Li Deren, Ma Jun, Shao Zhenfeng. Innovation in the census and monitoring of geographical national conditions [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(1): 1-9. (in Chinese)

[4] Yang Bisheng, Dong Zhen. Progress and perspective of point cloud intelligence [J]. *Acta Geodetica et Cartographica Sinica*, 2019, 48(12): 1575-1585. (in Chinese)

[5] Li Deren. From geomatics to geospatial intelligent service science [J]. *Acta Geodetica et Cartographica Sinica*, 2017, 46(10): 1207-1212. (in Chinese)

[6] Vosselman G, Mass H G. Airborne and terrestrial laser scanning[M]. Boca Ration, FL: CRC Press, 2010.

[7] Hu Han, Ding Yulin, Zhu Qing, et al. An adaptive surface filter for airborne laser scanning point clouds by means of regularization and bending energy [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 92: 98-111.

[8] Sui Lichun, Zhang Yibin, Liu Yan, et al. Filtering of airborne LiDAR point cloud data based on the adaptive mathematical morphology [J]. *Acta Geodetica et Cartographica Sinica*, 2010, 39(4): 390-396. (in Chinese)

[9] Sui Lichun, Yang Yun. Filtering of airborne LiDAR point cloud data based on car(p, q) model and mathematical morphology [J]. *Acta Geodetica et Cartographica Sinica*, 2012, 41(2): 219-224. (in Chinese)

[10] Zhu Xiaoxiao, Wang Cheng, Xi Xiaohuan, et al. Research

- progress of ICESat-2/ATLAS data processing and applications [J]. *Infrared and Laser Engineering*, 2020, 49(11): 20200259. (in Chinese)
- [11] Qin Wei, Chen Xi, Ma Yuanyuan, et al. The analysis of edge detection algorithms based on mathematical morphology [J]. *Information Technology*, 2019, 43(11): 33-36. (in Chinese)
- [12] Zhao Chuan, Zhang Baoming, Guo Haitao, et al. Airborne LiDAR point cloud classification using transfer learning [J]. *Optics and Precision Engineering*, 2019, 27(7): 1601-1612. (in Chinese)
- [13] Liu Zhiqing, Li Pengcheng, Chen Xiaowei, et al. Classification of airborne LiDAR point cloud data based on information vector machine [J]. *Optics and Precision Engineering*, 2016, 24(1): 210-219. (in Chinese)
- [14] Du Ruijian, Ge Baozhen, Chen Lei. Texture mapping of multi-view high-resolution images and binocular 3D point clouds [J]. *Chinese Optics*, 2020, 13(5): 1055-1064. (in Chinese)
- [15] Zong Wenpeng, Li Guangyun, Li Minglei, et al. A survey of laser scan matching methods [J]. *Chinese Optics*, 2018, 11(6): 914-930. (in Chinese)
- [16] Chen Ying, Zhu Ming, Li Zhaoze. Remote sensing digital image enhancement based on Gaussian mixture modeling [J]. *Chinese Journal of Lasers*, 2014, 41(12): 1209002. (in Chinese)
- [17] Tao Zhiyong, Liu Xiaofang, Wang Hezhang. Clustering algorithm of Gaussian mixture model based on density peaks [J]. *Journal of Computer Applications*, 2018, 38(12): 3433-3437, 3443. (in Chinese)
- [18] Feng Fajie, Ding Yazhou, Li Junping. Airborne LiDAR point cloud filtering using saliency division [J]. *Infrared and Laser Engineering*, 2020, 49(8): 20190439. (in Chinese)
- [19] Sun Hao, Guo Yingqing, Zhao Wanli. Improved model for on-board real-time by constructing empirical model via GMM clustering method [J]. *Journal of Northwestern Polytechnical University*, 2020, 38(3): 507-514. (in Chinese)
- [20] Sithole G, Vosselman G. A comparison of existing automatic filters[R]. Technical Report, ISPRS test on extracting DEMs from point clouds, 2003.