

基于多尺度自适应特征的机载 LiDAR 点云分类

杨书娟^{1,2}, 张珂殊^{2*}, 邵永社²

¹中国科学院大学, 北京 100049;

²中国科学院电子学研究所, 北京 100190

摘要 为解决复杂场景下城区点云分类精度不高的问题, 提出了基于多尺度自适应特征的分类方法。首先, 对经典几何统计特征和点直方图特征进行组合, 将组合特征集作为分类依据; 然后采用随机森林法评估特征的重要性, 并自适应选取重要的特征集; 最后基于多尺度自适应特征实现了点云的分类。实验结果表明: 该方法能实现城区点云的高精度分类, 能适合任意尺度下不同分辨率点云数据的分类。

关键词 遥感; 点特征直方图; 自适应特征; 多尺度; 分类

中图分类号 TN958.98

文献标识码 A

doi: 10.3788/AOS201939.0228001

Classification of Airborne LiDAR Point Cloud Data Based on Multiscale Adaptive Features

Yang Shujuan^{1,2}, Zhang Keshu^{2*}, Shao Yongshe²

¹ University of Chinese Academy of Sciences, Beijing 100049, China;

² Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China

Abstract To solve the low-classification accuracy problems of urban point clouds in complex environments, we propose a classification method based on multiscale adaptive features herein. Firstly, the classical geometric statistical features and point histogram features are combined; then, the combined feature set is used for classification basis. Random forest is then used to assess the importance of the features and adaptively select important feature sets. Finally, the point clouds are classified based on these multiscale adaptive features. Experimental results reveal that this method can achieve a high-accuracy classification for point clouds in urban areas. The proposed method can be applied to the classification of point cloud data with different resolutions at arbitrary scale.

Key words remote sensing; histogram of point feature; adaptive features; multiscale; classification

OCIS codes 280.3420; 280.3640; 280.5600

1 引 言

激光雷达(LiDAR)技术作为一种新的技术手段被广泛应用于城市规划、应急响应和城市三维建模等方面,而点云的高精度分类是实现这些应用的前提^[1-2]。深入挖掘点云分类算法一直都是点云领域研究的热点问题,如:隋立春等^[3]、孙美玲等^[4]采用传统点云分类方法,基于滤波、高程阈值和梯度阈值等,实现了点云分类,但该方法需设置较多的阈值,自动化程度较低。有学者试图从 LiDAR 数据特征描述和智能学习方面来提高 LiDAR 地物的分类精度,如:乔纪刚等^[5]、杨应等^[6]、Tiwari 等^[7]、王

晓玲等^[8]将点云内插成高程纹理影像,采用机器学习进行点云分类,但将高分辨率点云数据转换成栅格图像时,会造成不可逆的信息丢失。因此,直接利用 LiDAR 数据的三维特性进行地物分类是近年研究的热点,如:范士俊等^[9]、Narayanan 等^[10]、Lodha 等^[11]、Shu 等^[12]、Zhang 等^[13]将激光点云的几何特征(高程、法向量)和波形特征(振幅、回波次数和波宽)进行融合,采用随机森林、支持向量、AdaBoost 集成等机器学习的方法实现点云分类。但以上方法都忽略了冗余特征以及分类能力较差的特征对分类精度和效率的影响,因此,张爱武等^[14]、姚登举等^[15]、孙杰等^[16]提出要先对点云特征的重要性进

收稿日期: 2018-05-21; 修回日期: 2018-06-10; 录用日期: 2018-06-17

* E-mail: ks Zhang@mail.ie.ac.cn

行分析,再基于最优特征集实现点云分类,然而该方法未考虑尺度的选择对分类精度的影响。

考虑到不同尺度特征、冗余特征、分类能力较差的特征对分类精度的影响,本文提出了基于不同尺度自适应特征的分类方法。该方法首先对经典几何统计特征和点直方图特征进行组合,将组合特征集作为分类依据;然后采用随机森林法评估特征的重要性,并对特征进行排名,去除能力较弱的特征,自适应选取重要的特征集;最后基于多尺度自适应特征,实现 LiDAR 点云的分类。

2 LiDAR 数据的特征描述算子

经典几何统计特征包含的信息量较少,对点类别的区分度低,无法实现城市区域的精确分类。点特征直方图(PFH)通过构造高维空间表达点的几何特征来体现邻域的几何分布特征。因此,本研究将经典几何统计特征和 PFH 特征相组合,同时考虑了点邻域的统计特征和分布特征。

2.1 经典的几何统计特征

原始激光点云数据不能直接用于智能学习机分类,需要经过数据变换、特征提取和训练等环节后才能输入分类器。点云的几何特征可以通过对该点的三维邻域内所有点进行统计分析得到,主要包括:1)与高程相关的特征,即高程均方差(HD)和高程极差(NH);2)与投影相关的特征——投影后的点密度(pointDensity);3)与面相关的特征,即法向量竖直角 $\left[\text{CON}, \theta = a \tan \left(\frac{n_z}{\sqrt{n_x^2 + n_y^2}} \right) \right]$ 、法向量竖直角的均值(angleAve)、法向量竖直角的均方差(angleStd)、平面拟合残差均值(AveDist);4)与协方差矩阵相关的特征,即平整性 $\left\{ \text{EVE}, \delta = \left[\frac{\lambda_0}{(\lambda_0 + \lambda_1 + \lambda_2)} \right] \right\}$ 、分布性 $\left(\text{DIS}, \varepsilon = \sqrt{\prod_{i=0}^2 \lambda_i} \right)$ 、线性特征 $\left[L, \eta = \frac{(\lambda_1 - \lambda_2)}{\lambda_1} \right]$ 、平面指数 $\left[P, \tau = \frac{(\lambda_2 - \lambda_3)}{\lambda_1} \right]$ 、球面指数 $\left[S, \kappa = \frac{\lambda_3}{\lambda_1} \right]$ 、信息熵 $\left[E, \mu = \sum_{i=0}^2 \lambda_i \ln(\lambda_i) \right]$ 和高斯曲率 $\left(V_{\text{ce}}, \chi = \sqrt{\prod_{i=0}^2 \lambda_i} \right)$,其中, $(\lambda_0, \lambda_1, \lambda_2)$ 为邻域内点云构成协方差矩阵的特征值, (n_x, n_y, n_z) 为最小特征值 λ_0 对应的特征向量。

2.2 PFH 特征

PFH 通过构造高维空间表达点的几何特征来更加详细地表述中心点与邻域点之间的关系。为了描述点与其邻域点之间的空间关系,需要建立一个

局部坐标系,如图 1 所示。

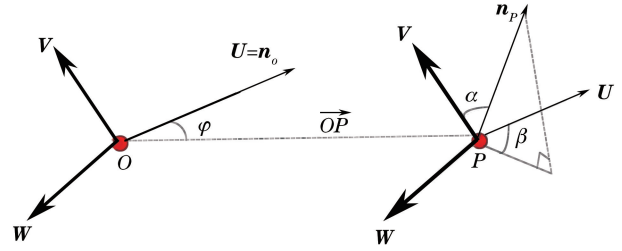


图 1 相邻点的局部空间坐标系

Fig. 1 Local space coordinate system of adjacent points

以 O 为原点建立局部坐标系, U 与 O 的法向量 n_o 平行, V 垂直于 U 与 \overrightarrow{OP} 构成的平面, W 垂直于 U 与 V 构成的平面, n_p 为 O 点邻域点 P 的法向量。U、V、W 的定义为

$$\begin{cases} U = n_o \\ V = U \times \frac{\overrightarrow{OP}}{|\overrightarrow{OP}|} \\ W = U \times V \end{cases} \quad (1)$$

因此,相邻两点法向量的空间关系可用三个角度参数 α, φ, β 表示,其中: α 为向量 n_p 与 V 的夹角, $\alpha = \arccos(\mathbf{V} \cdot \mathbf{n}_p)$; φ 为向量 n_o 与 \overrightarrow{OP} 的夹角, $\varphi = \arccos \left(\mathbf{U} \cdot \frac{\overrightarrow{OP}}{|\overrightarrow{OP}|} \right)$; β 为向量 n_p 在 WPU 平面上的投影与坐标轴 U 的夹角, $\beta = \arctan[\mathbf{W} \cdot \mathbf{n}_p / (\mathbf{U} \cdot \mathbf{n}_p)]$ 。

PFH 采用三参数量化中心点与邻域点之间的空间关系,形成一个用于描述中心点邻域几何属性的多维直方图。假设某激光点的邻域内有 n 个激光点,计算邻域内所有两点间的三参数,可得到 C_n^3 个三元素,把三个角度分别划分为 b 个子区间,统计邻域内 C_n^3 个三元素落在子区间的点数目,即可得到 $3 \times b$ 个子区间的直方图。

3 基于多尺度自适应特征的分类

3.1 多尺度特征

尺度的选择会直接影响点云的分类精度,因此需要根据场景内的地物选择合适的尺度。通常,不同类别的地物在多种尺度下特征的区别越明显,说明地物的可分性越强。例如:地面上有灌木丛时,在大尺度下(0.5 m),地面点被误分类为植被点,而在小尺度下(0.2 m),地面点会被正确分类,如图 2 所示;然而,建筑物边缘在大尺度下才能被正确分类,如图 3 所示。

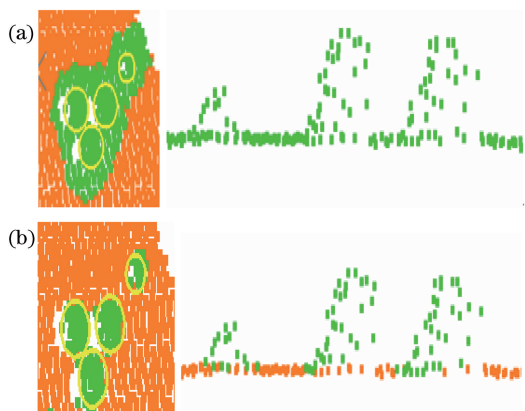


图 2 灌木丛区域的分类。(a)大尺度；(b)小尺度

Fig. 2 Classification of bush area.

(a) Large scale; (b) small scale

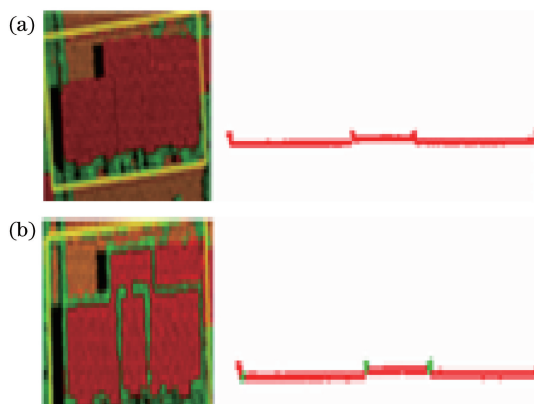


图 3 建筑物的分类。(a)大尺度；(b)小尺度

Fig. 3 Classification of building.

(a) Large scale; (b) small scale

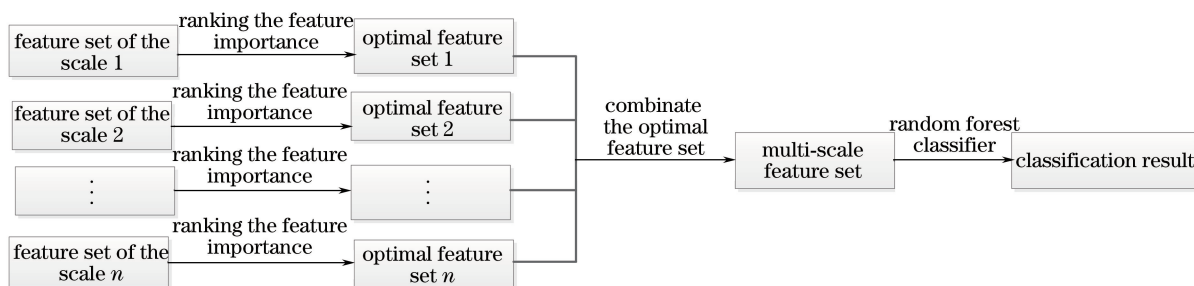


图 4 自适应多尺度特征分类的流程图

Fig. 4 Flowchart of multiscale adaptive-feature classification

计算多尺度特征比计算单尺度特征更耗时。为了减少处理时间,对于大场景数据,可根据地物类型对大场景进行分块,选择合适的尺度(小尺度、大尺度或多尺度)进行分类。如:建筑物区域适合采用大尺度分类,植被区域适合采用小尺度分类,而同时存在建筑物、植被和地面等的地物复杂区域则适合使用多尺度分类。

对于环境复杂的城区,选择合适的邻域半径尤为重要。因此,本研究将 LiDAR 数据的多尺度几何特征作为分类依据,该方法能适应任意尺度下不同密度的 LiDAR 数据分类,且能够应用于大数据和大自然场景的点云数据分类。

3.2 多尺度自适应特征分类

分类能力较弱的特征和特征集合中的冗余特征会增加分类的复杂性,降低点云分类的效率和精度,本研究采用随机森林变量重要性算法评估特征的分类能力,依据重要程度系数的高低对特征进行排序,去除排名较低的特征,得到适合点云分类的最优特征集合。

图 4 为多尺度自适应特征分类的流程图,具体流程如下:

1)设置尺度范围及尺度间隔。最小尺度一般大于点云的最小密度,最大尺度可根据建筑物的平均宽度定义,按最小尺度依次增加尺度间隔的大小。

2)对于某一个尺度下的特征集合 F ,利用随机森林特征重要性算法,计算 F 中每个特征的分类重要程度系数,根据特征分类重要程度系数的高低,依次将特征加入到特征集合中,并统计分类精度。当分类精度达到最高时,设定特征分类重要程度系数阈值 I ,去除特征集 F 中排名靠后的特征,得到最优特征集合 F' 。

3)重复步骤 2),得到不同尺度的最优特征集合。

4)将不同尺度的最优特征集合进行合并,再基于随机森林分类器,实现 LiDAR 的分类。

4 实验与分析

实验数据由天绘机载激光扫描仪获取。该机载激光扫描仪的平均飞行高度为 500 m,平均点间距为 0.1 m,飞行区域为城市区域,该区域分布有建筑物、灌木丛和道路。图 5 为实验数据按高程显示的效果图。

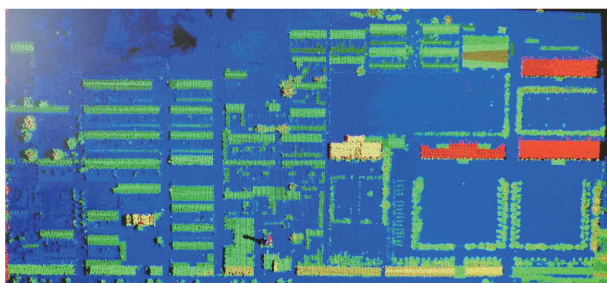


图 5 实验数据

Fig. 5 Experimental data

4.1 基于不同特征组合的分类结果

利用随机森林特征变量重要性算法对三组组合的特征重要程度进行分析,并进行降序排列,结果如

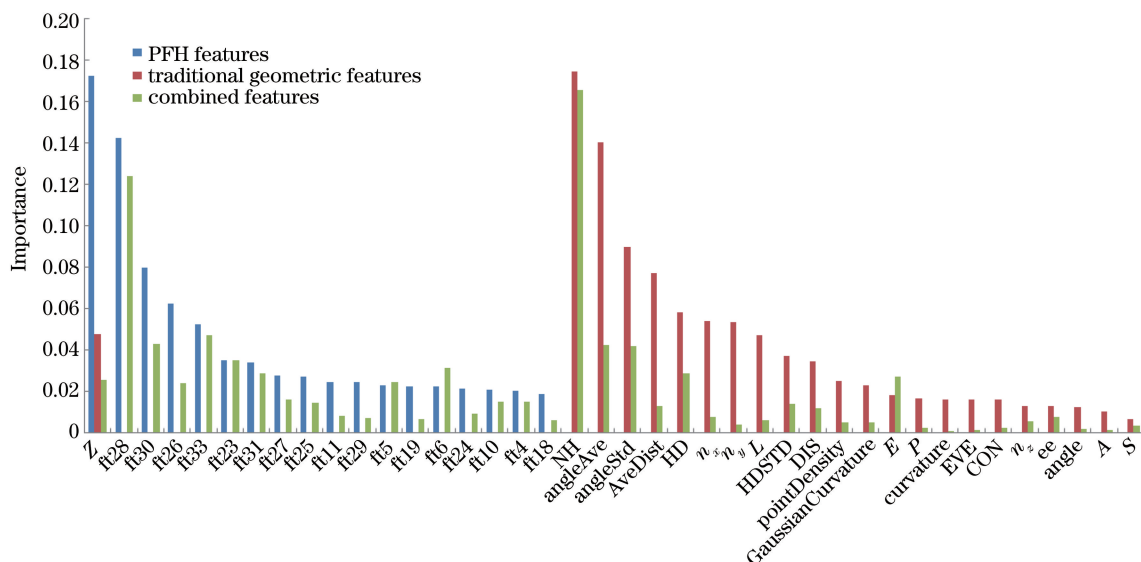


图 6 不同特征组合的特征重要性分布

Fig. 6 Distribution of feature importance for different feature combinations

针对三种组合的特征集,去除分类能力较弱的特征和特征集合中的冗余特征后,采用随机森林法实现点云分类,分类结果如图 7 所示。可以看到:1)对于经典几何统计特征集的分类结果,由于建筑物边缘点的邻域表面特征、高程特征、法向量特征与树木的邻域特征比较接近,故而,建筑物边缘点被误分类为植物,但 3 m 高的树木被正确分类为植物,如图 7(a)所示;2)对于 PFH 特征集的分类结果,建筑物边缘被正确分类为建筑物,但部分 3 m 高的树木的邻域 PFH 特征与建筑物的邻域 PFH 特征相似,因此树木被误分类为建筑物,如图 7(b)所示;3)对于组合特征集的分类结果,建筑物边缘和 3 m 高的树木都能被正确分类,如图 7(c)所示。

用验证样本计算不同特征集的分类精度,结果如表 1 所示。可以看到:对于地面而言,三种类型特征集的分类精度基本一致;对于植被而言,经典几何

图 6 所示。可以看到:1)对于经典几何统计特征集,高程特征 NH、HD 和角度特征 angleAve、angleStd、 n_x 、 n_y 、AveDist 的重要性较高,而协方差相关特征 (E、P、curvature、EVE、CON、ee、A、S)的重要性较小;2)对于 PFH 特征集,高程值 Z、ft28、ft30、ft26、ft33、ft23、和 ft31 的重要性较高;3)对于组合特征集,高程特征 NH、HD、角度特征 angleAve、angleStd、高程值 Z、ft28、ft30、ft26、ft33、ft23 和 ft31 的重要性较高。组合特征集的重要特征集相当于经典几何统计特征集和 PFH 特征集的重要特征集的组合,说明两组特征的相关性小,因此,组合后的特征集能更好地区分不同的地物,从而提高分类精度。

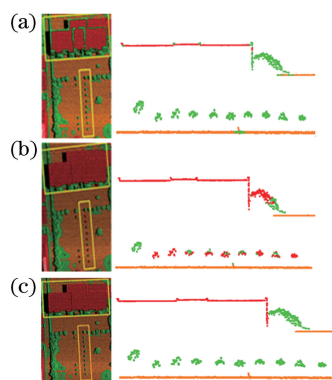


图 7 不同特征集的分类结果。(a)经典几何统计特征集; (b) PFH 特征集; (c)组合特征集

Fig. 7 Classification results of different feature sets.

(a) Classical geometric statistical feature set; (b) PFH feature set; (c) combinational feature set

统计特征集和组合特征集的分类精度较高;对于建筑物而言,PFH 特征集和组合特征集的分类精度较

表 1 不同类型特征集的分类精度

Table 1 Classification accuracy for different feature sets

Category	Classification of classical geometric statistical feature set / %	Classification of PFH feature set / %	Classification of combinational feature set / %
Ground	93.26	93.32	93.43
Vegetable	91.52	90.12	91.65
Building	90.64	92.43	92.72

高。可见,基于组合特征集的分类精度更高。

4.2 基于多尺度自适应特征的分类结果

对于环境复杂的区域,选择合适的邻域半径尤为重要,因此,本研究采用多尺度特征进行分类。实验数据的平均点密度为 0.1 m,最小尺度选为 0.1 m,因此,根据建筑物的平均宽度,最大尺度设置为 1 m,考虑到分类精度和时间,尺度间隔设置为

2 倍的点间距(0.2 m),根据尺度间隔划分为 5 个尺度等级。

针对组合特征集,分别计算 5 组尺度下的特征重要性,结果如图 8 所示。可以看到,不同邻域半径会直接影响特征的重要性分布,即不同尺度的相同特征对分类结果的影响也不同,因此,多尺度下的特征能更好地区分不同的地物,提高分类精度。

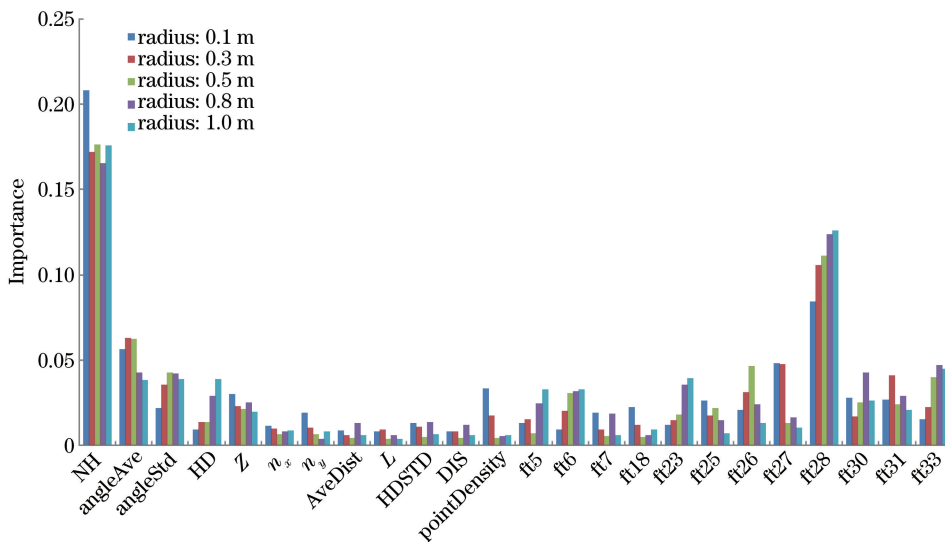


图 8 不同尺度下的特征重要性分布

Fig. 8 Distribution of feature importance at different scales

将不同尺度的重要特征集组合后,采用随机森林分类器实现城区点云的分类,如图 9 所示。图中绿色为植被,红色为建筑物,橙色为地面。可以看到:1)基于大尺度特征集(邻域半径为 1 m)分类时,E 区域的地面基本被正确分类,但 A 区域植被周围的地面被误分类为植被,B 区域狭长的建筑物被误分类为植被;2)基于小尺度特征集(邻域半径为 0.1 m)分类时,B 区域狭长的建筑物被正确分类,但 A 区域大量的地面点被误分类为植被,E 区域的少量植被和地面被误分类为建筑物;3)基于多尺度特征集(邻域半径为 0.1~1 m)分类时,A、B 和 E 区域的点云都被正确分类。

对于 C 区域(建筑物和树木)而言,不同尺度特征集的分类结果如图 10 所示。可以看出:1)基于大尺度特征集(邻域半径为 1 m)分类时,建筑物边缘

和屋顶凸起物,以及 3 m 高的树木被正确分类,但树木附近的地面点被误分类为植被,且邻域半径越大,被误分类的地面点越多,如图 10(a)所示;2)基于小尺度特征集(邻域半径为 0.1 m)分类时,低矮树木附近的地面点被正确分类为地面点,但建筑物边缘和屋顶凸起物被误分类为植被,如图 9(b)所示;3)基于多尺度特征集分类时,3 m 高的树木和树木附近的地面点、建筑物边缘和屋顶凸起物都能被正确分类,如图 10(c)所示。

对于 D 区域(相邻房屋的狭窄马路)而言,不同尺度特征集的分类结果如图 11 所示。可以看到:1)基于大尺度特征集(邻域半径为 1 m)分类时,建筑物被正确分类,但相邻房屋的狭窄马路(黄色框区域)被误分类为植被,如图 11(a)所示;2)基于小尺度特征集(邻域半径为 0.1 m)分类时,相邻房屋的

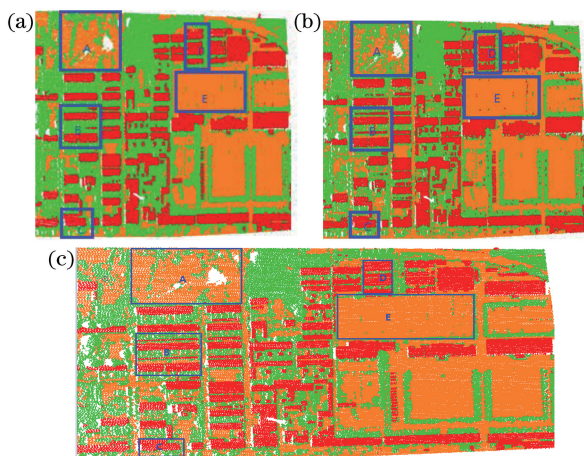


图 9 基于不同尺度特征集的分类结果。

(a)大尺度; (b)小尺度; (c)多尺度

Fig. 9 Classification results based on feature set with different scales. (a) Large scale; (b) small scale; (c) multiscale

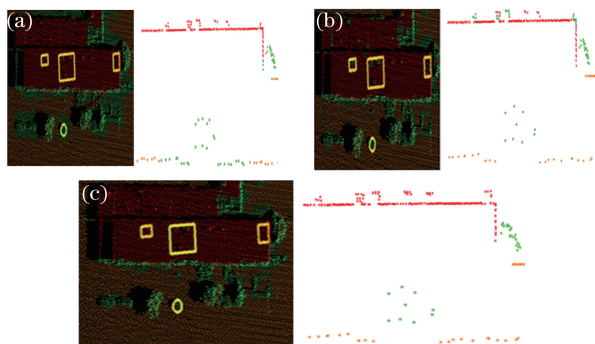


图 10 C 区域基于不同尺度特征集的分类结果。

(a)大尺度; (b)小尺度; (c)多尺度

Fig. 10 Classification results of C area based on feature set with different scales. (a) Large scale; (b) small scale; (c) multiscale

狭窄马路被正确分类为地面点,但建筑物凸起被错误分类为植被,如图 11(b)所示;3)基于多尺度特征集分类时,相邻房屋的狭窄马路和建筑物凸起都能被正确分类,如图 11(c)所示。

用验证样本计算不同尺度的分类精度,结果如表 2 所示。可以看到:随着尺度增大,地面和植被的分类精度先增加后降低,而建筑物的分类精度则逐渐增大;对于多尺度而言,三类地物点云的分类精度均有所提高。

5 结 论

本课题组研究了城区点云的多尺度组合特征,采用随机森林分类器实现了城区地面、植被和建筑物 LiDAR 数据的高精度分类。实验结果表明:1)组

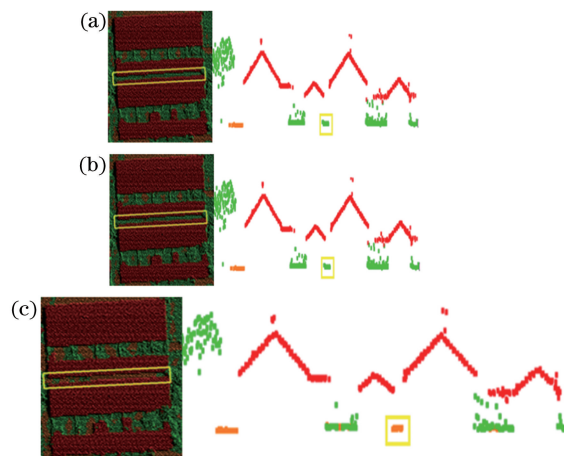


图 11 D 区域基于不同尺度特征集的分类结果。

(a)大尺度; (b)小尺度; (c)多尺度

Fig. 11 Classification results of D area based on feature set with different scales. (a) Large scale; (b) small scale; (c) multiscale

表 2 不同尺度下的分类精度

Table 2 Classification accuracy at different scales

Scale /m	Classification accuracy /%		
	Ground	Vegetable	Building
0.1-1.0	95.83	95.36	94.86
0.1	85.33	86.29	79.42
0.3	92.64	94.75	82.32
0.5	86.72	89.63	85.94
0.8	82.53	87.49	89.87
1.0	78.68	82.63	92.12

合特征集的分类精度高于单一特征集(经典几何统计特征集或 PFH)的分类精度;2)多尺度特征集的分类精度高于单一尺度特征集的分类精度,该方法避免了尺度选择问题,能适应不同点分辨率点云的高精度分类。

参 考 文 献

- [1] Zhang L M. The method of airborne laser scanning data classification[D]. Chengdu: Southwest Jiaotong University, 2013.
张利明. 机载激光雷达点云数据分类方法研究[D]. 成都: 西南交通大学, 2013.
- [2] Chen G B, Gao Z H, He L. Step-by-step automatic calibration algorithm for exterior parameters of 3D LiDAR mounted on vehicle[J]. Chinese Journal of Lasers, 2017, 44(10): 1010004.
陈贵宾, 高振海, 何磊. 车载三维激光雷达外参数的分步自动标定算法[J]. 中国激光, 2017, 44(10): 1010004.
- [3] Sui L C, Zhang Y B, Liu Y, et al. Filtering of airborne LiDAR point cloud data based on the adaptive

- mathematical morphology [J]. *Acta Geodaetica et Cartographica Sinica*, 2010, 39(4): 390-396.
- 隋立春, 张熠斌, 柳艳, 等. 基于改进的数学形态学算法的 LiDAR 点云数据滤波 [J]. *测绘学报*, 2010, 39(4): 390-396.
- [4] Sun M L, Li Y S, Chen Q, *et al.* Iterative multi-scale filter based on morphological opening by reconstruction for LiDAR urban data [J]. *Infrared and Laser Engineering*, 2015, 44(1): 363-369.
- 孙美玲, 李永树, 陈强, 等. 基于迭代多尺度形态学开重建的城区 LiDAR 滤波方法 [J]. *红外与激光工程*, 2015, 44(1): 363-369.
- [5] Qiao J G, Liu X P, Zhang Y H. Land cover classification using LiDAR height texture and ANNs [J]. *Journal of Remote Sensing*, 2011, 15(3): 539-553.
- 乔纪纲, 刘小平, 张亦汉. 基于 LiDAR 高度纹理和神经网络的 地物分类 [J]. *遥感学报*, 2011, 15(3): 539-553.
- [6] Yang Y, Su G Z, Zhou M. Filtering of LiDAR point clouds data with image classification information [J]. *Geomatics and Information Science of Wuhan University*, 2010, 35(12): 1453-1456.
- 杨应, 苏国中, 周梅. 影像分类信息支持的 LiDAR 点云数据滤波方法研究 [J]. *武汉大学学报(信息科学版)*, 2010, 35(12): 1453-1456.
- [7] Tiwari P S, Pande H. Use of laser range and height texture cues for building identification [J]. *Journal of the Indian Society of Remote Sensing*, 2008, 36(3): 227-234.
- [8] Wang X L, Du P J. Spatial and spectral classification based on morphology using SVM [J]. *Bulletin of Surveying and Mapping*, 2012(12): 18-22.
- 王晓玲, 杜培军. 基于形态学的空间信息和光谱信息 SVM 影像分类 [J]. *测绘通报*, 2012(12): 18-22.
- [9] Fan S J, Zhang A W, Hu S X, *et al.* A method of classification for airborne full waveform LiDAR data based on random forest [J]. *Chinese Journal of Lasers*, 2013, 40(9): 0914001.
- 范士俊, 张爱武, 胡少兴, 等. 基于随机森林的机载激光全波形点云数据分类方法 [J]. *中国激光*, 2013, 40(9): 0914001.
- [10] Narayanan R, Kim H B, Sohn G. Classification of SHOALS 3000 bathymetric LiDAR signals using decision tree and ensemble techniques [C]. 2009 IEEE Toronto International Conference Science and Technology for Humanity, Toronto, 2009: 462-467.
- [11] Lodha S K, Fitzpatrick D M, Helmbold D P. Aerial lidar data classification using AdaBoost [C]. Sixth International Conference on 3-D Digital Imaging and Modeling, Montreal, 2007: 435-442.
- [12] Shu Z, Sun K, Qiu K J, *et al.* Pairwise-SVM for on-board urban road lidar classification [J]. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016(8): 109-113.
- [13] Zhang J X, Lin X G, Ning X G. SVM-based classification of segmented airborne LiDAR point clouds in urban areas [J]. *Remote Sensing*, 2013, 5(8): 3749-3775.
- [14] Zhang A W, Xiao T, Duan Y H. A method of adaptive feature selection for airborne LiDAR point cloud classification [J]. *Laser & Optoelectronics Progress*, 2016, 53(8): 082802.
- 张爱武, 肖涛, 段乙好. 一种机载 LiDAR 点云分类的自适应特征选择方法 [J]. *激光与光电子学进展*, 2016, 53(8): 082802.
- [15] Yao D J, Yang J, Zhan X J. Feature selection algorithm based on random forest [J]. *Journal of Jilin University (Engineering and Technology Edition)*, 2014, 44(1): 137-141.
- 姚登举, 杨静, 詹晓娟. 基于随机森林的特征选择算法 [J]. *吉林大学学报(工学版)*, 2014, 44(1): 137-141.
- [16] Sun J, Lai Z L. Airborne LiDAR feature selection for urban classification using random forests [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(11): 1310-1313.
- 孙杰, 赖祖龙. 利用随机森林的城区机载 LiDAR 数据特征选择与分类 [J]. *武汉大学学报(信息科学版)*, 2014, 39(11): 1310-1313.