

综合布料滤波与改进随机森林的点云分类算法

薛豆豆^{1*}, 程英蕾¹, 释小松¹, 秦先详¹, 文沛^{1,2}

¹空军工程大学信息与导航学院, 陕西 西安 710077;

²93575 部队, 河北 承德, 067000

摘要 城区激光雷达点云建筑物提取技术是近年来发展的热点, 如何准确区分植被、建筑物以及人造物, 提高分类精度一直是研究难点。为此, 针对分类精度较低的问题, 提出一种基于随机森林的点云分类算法。首先使用改进布料滤波算法对点云数据进行地面滤波; 其次, 构建决策树并进行基于最大互信息系数的相关性分析, 选出相关系数最小、精度最高的决策树, 得到弱相关随机森林模型; 最后, 对决策结果进行加权投票处理, 得到一种综合布料滤波和加权弱相关随机森林的点云分类算法, 并通过 Vaihingen 城区数据集对算法进行验证。实验表明, 与传统随机森林分类算法相比, 本文算法提高了 4.2% 的分类精度, 也提高了算法效率。

关键词 图像处理; 激光雷达; 布料滤波算法; 随机森林; 点云归一化; 最大互信息系数

中图分类号 P237

文献标志码 A

doi: 10.3788/LOP57.221017

Point Clouds Classification Algorithm Based on Cloth Filtering Algorithm and Improved Random Forest

Xue Doudou^{1*}, Cheng Yinglei¹, Shi Xiaosong¹, Qin Xianxiang¹, Wen Pei^{1,2}

¹Information and Navigation College, Air Force Engineering University, Xi'an, Shaanxi 710077, China;

²The 93575 Unit, Chengde, Hebei 067000, China

Abstract Building extraction technology in urban areas has been a hot topic in recent years, but how to accurately distinguish vegetation, buildings, and man-made objects and improve classification accuracy has always been a difficult point. Aiming at the problem of low classification accuracy, we propose a point cloud classification algorithm based on random forest. First, the improved cloth filtering algorithm is used to perform ground filtering on the point cloud data. And a decision tree is constructed and the correlation analysis based on the largest mutual information coefficient is performed to select the decision tree with the smallest correlation coefficient and the highest accuracy to obtain a weakly correlated random forest model. The decision results are processed by weighted voting, and finally a point cloud classification algorithm combining cloth filtering and weighted weakly correlated random forest is obtained. Compared with the traditional random forest classification algorithm, the algorithm is verified by the Vaihingen urban dataset, and the classification accuracy is improved by 4.2%.

Key words image processing; LiDAR; cloth filtering algorithm; random forest; normalized point cloud; maximal information coefficient

OCIS codes 100.4996; 280.3420; 280.3640

1 引言

近年来, 随着城市化建设的不断加快, 城市人口过多, 地面资源短缺的问题日益严重。城市的快速发展对建设“智慧城市”的需求日益紧迫。其中, 快

速准确地获取城市对象信息的能力是必不可少的。激光雷达(LiDAR)技术凭借其高效、高分辨率获取地形信息且不受天气影响的特点, 已经被广泛应用于城市目标分类以及建筑物提取等多项研究中, 目前点云数据的精确分类一直是研究的热点和难点。

收稿日期: 2020-02-01; 修回日期: 2020-02-27; 录用日期: 2020-03-06

基金项目: 国家自然科学基金(41601436)

* E-mail: 1447551957@qq.com

点云分类算法主要有两种类型:同时分类算法和阶层式分类算法^[1]。同时分类方法直接利用点云特征对各个物体分类,文献[2]提出了引入条件随机场模型对点云数据进行分类,文献[3]是利用条件随机场和随机森林(RF)相结合的方式实现点云的分类;同时分类算法由于同时处理大量点云数据,造成算法效率低。阶层式分类算法分为地面滤波和对对象提取两个步骤。传统的滤波方法有坡度滤波^[4]、形态学滤波、不规则三角网滤波、曲面滤波等。Vosselman等^[5]通过利用坡度值差异构建滤波算法,Kilian等^[6]提出形态学滤波。Sithole等^[7]对8种传统滤波算法进行了实验比较,多数滤波算法需要用户自定义多个参数,导致算法滤波效果不稳定。Zhang等^[8]提出了一种基于布料模拟(CSF)的滤波算法,通过使用布料对地面进行拟合,算法分类精度较高,滤波效果稳定。本文在保证算法精度情况下对布料的网格化进行改进,简化了算法步骤。在对象提取方面,需要对不同类型的特征进行分析,由于机器学习算法可以自动定义分类规则而被广泛应用其中。文献[9]首先使用地面滤波算法分离地面点,接着根据建筑物形态特征进行地物点分离。Mallet等^[10]利用支持向量机(SVM)算法进行城区点云分类,Guo等^[11]采用随机森林算法对点云进行分类,取得良好的效果。Coussement等^[12]通过对SVM、随机森林以及Logistic模型对比,发现随机森林模型的预测能力最好,优于其他两种模型。为了进一步提高模型性能,Battiti等^[13]通过计算特征之间的最小冗余提出了互信息特征选择算法,文献[14]进一步通过优化评价函数等方法提高随机森林算法分类精度。但是上述算法只考虑了特征之间的相关性,并未对决策树的相关性进行考虑。

针对上述问题,本文提出了一种综合布料滤波和加权弱相关随机森林的点云分类算法,对城区点云进行滤波及分类。首先使用改进布料滤波(ICSF)算法对点云进行过滤,分离出地面点;接着根据最大互信息系数(MIC)矩阵和样本相关性计算对决策树之间的相关性进行评估;最后根据决策树精度对各个决策树进行加权,构建加权弱相关随机森林,实现了点云数据的精确分类。

2 改进布料滤波算法

本文采用的是阶层式分类算法,需要首先对点云进行滤波。本文采用ICSF算法,其基本原理是:首先通过对点云的高程差与阈值比较去除噪声;接

着为了消除地形起伏对滤波的影响,对点云进行归一化处理,再将点云翻转,使模拟布料置于点云上方;布料受到重力落到地形表面,初步判定布料首先接触的的点为地面点;然后分析布料点的重力以及点之间的内部作用力,确定布料的形状;最后通过将布料点与原始点云距离与阈值对比,完成地面点提取。算法流程图如图1所示。

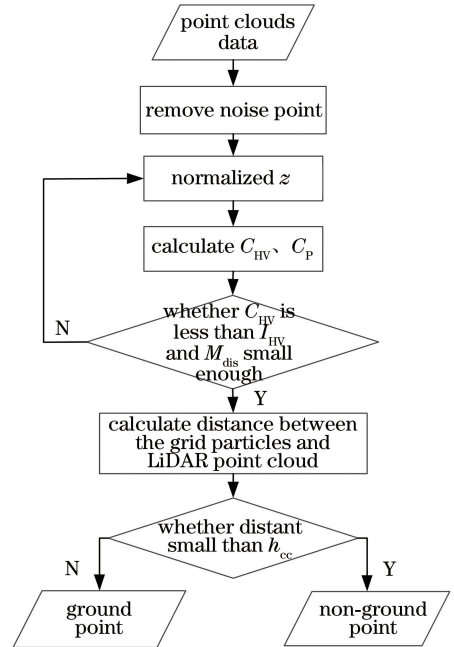


图1 ICSF滤波算法流程图

Fig. 1 Flow chart of ICSF filtering algorithm

ICSF算法通过加入重力条件,使拟合点能够初步确定地面点的位置,接着通过内部作用力对布料点位置进一步约束,完成滤波。具体步骤如下:

1)点云去噪。通过高斯滤波算法计算点云高程的均值以及标准差,将距离在标准差范围之外的点从数据集剔除,实现点云去噪。

2)点云归一化。将每个点 P 的原始高程值 z 与邻域高程均值相减,得到归一化高程值 z' 。

3)布料网格化。将预处理的点云翻转,将所有的点云投影到二维平面,平面上所有的点作为格网点,其对应的原始点云作为 C_P ,记录每个点的高程值为 I_{HV} 。

4)迭代移动过程。遍历所有的网格点,计算该点由重力移动的距离 C_{HV} ,

$$C_{HV} = 2X(t) - X(t - \Delta t) + \frac{G}{m} \Delta t^2, \quad (1)$$

式中: $X(\cdot)$ 为布料在某一时刻的位置; Δt 为时间间隔; G 为重力系数; m 为布料点质量,本文中取值为1。当小于或等于临界值 I_{HV} 时,移动格网点到

对应点云的位置,并标记为不可动点。接着计算相邻布料点高程差,计算在内部作用力下布料点的位移 d_{dis} 。重复此过程,直到所有布料点的最大高程位移值 M_{dis} 小于阈值,迭代过程结束。

$$d_{\text{dis}} = \frac{1}{2}n(p_j - p) \cdot \mathbf{Z} \cdot f, \quad (2)$$

式中: n 表示是否已经标记为不动点,若已经标记为不动点,则 $n=0$,否则, $n=1$; p 为当前节点位置; p_j 为临近节点位置; \mathbf{Z} 为单位向量(0,0,1); f 为刚性系数。

5)地物点分类。计算每个网格点高程值与对应 I_{HV} 的距离。对于点云数据,如果小于 h_{cc} ,则判定为地面点,否则,为非地面点。重复此过程,直到所有点分类完毕。

ICSF 算法相比传统的 CSF 算法有两点改进:1)传统 CSF 在点云进行去噪后直接对点云进行处理,而 ICSF 会进一步对点云进行归一化处理,减轻地面高程对滤波精度的影响;2)传统 CSF 算法需要考虑点云的密度进而确定布料格网的大小,而 ICSF 算法可以自适应点云密度,在不降低算法精度的情况下,减少算法复杂度,简化了算法流程,减少了算法运行时间。

3 点云特征构建

在对点云进行滤波后,可以得到地物点。为了提取随机森林模型分类所需要的特征,本文主要从两个方面对激光 LiDAR 点云数据的特征进行整理:基本特征和邻域特征。基本特征主要反映独立点的特征,没有对局部的特征进行考虑,而邻域特征正是为了进一步完善特征集而提出的。

3.1 基本特征

基本特征主要是点云数据本身含有的特征,可以直接从设备中获取,主要包括点云的三维坐标、回波次数、全球定位系统(GPS)记录时间、强度等信息。本文用到的基本特征及含义说明如下:

1)回波次数 N_{Return} 。激光 LiDAR 在地面等平坦位置生成一次回波,而在植被等区域形成多次回波。

2)回波强度 I 。不同的介质表面对激光的反射特性不同。

3)归一化高程值 n_{Ele} 。在利用改进布料滤波算法对原始点云数据滤波后,通过将地物点减去距离该点最近的地面点高程值,得到归一化高程值 n_{Ele} 。归一化高程值可以有效分离高大建筑物与低矮植被。

3.2 邻域特征

邻域特征是对点云局部信息的提取。相关学者最先提出利用邻域的高程均值、高程方差、强度均值、高程变化范围等特征对点云进行分类。2014年,张蕊等^[15]在分类点云数据时对特征值进行主成分分析(PCA),达到了预期效果。本文通过对点云数据进一步分析,提出了法向量方差、拟合指数、线指数等特征。本文用到的邻域特征及含义说明如下:

1)高程均值 \bar{h} 、高程变化范围 ϵ_h 和高程方差 σ_h 。高程方差体现了物体表面的平整度。植被表面冠层不规则,高程方差较大;建筑物表面平整,高程方差小;同理可以得到点云回波强度的均值以及方差。

2)法向量方差 σ_θ 。通过拟合各点球形邻域内的点集合 P ,求拟合平面法向量与垂直方向夹角,而法向量方差是中心点与所有邻域点法向量夹角的方差。屋顶等平面方差变化较小,非平面如树木等法向量夹角方差变化较大。

3)拟合指数 μ_p 。拟合指数是指点云数据集中任意一点 p 邻域内各点到拟合平面的最小二乘距离均值。建筑物拟合指数相对于树木较小。

4)点云密度。计算单位体积内建筑物、房屋等点云密度特征。

5)邻域的最小特征值 λ_3 。点云的最小特征值对应的特征向量为点云拟合平面的法向量。首先计算点 p 邻域点集合 P' 的协方差矩阵 Σ ,并通过奇异值分解得到矩阵的特征值 λ 和特征向量 β 。假设矩阵 Σ 的特征值 $\lambda_1 > \lambda_2 > \lambda_3$ 以及特征向量 $\beta_1, \beta_2, \beta_3$,为了提高数据运算效率,将特征值进行归一化和缩放,公式为

$$\lambda'_i = \frac{(\lambda_i - \mu_i)}{\max(\lambda_i)}, i = 1, 2, 3, \quad (3)$$

式中: λ_i 为特征值; μ_i 为特征值的均值。

6)线指数 λ_1 和各向异性 σ_f 。通过点云各向异性和线指数等特征可以识别平面区域以及部分轮廓线。点云数据集中任意一点 p 的各向异性和线指数表示为

$$\lambda_1 = \frac{\lambda_1 - \lambda_2}{\lambda_1}, \quad (4)$$

$$\sigma_f = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3}. \quad (5)$$

上文从点云的基本特征和局部特征两方面对点云数据的特征进行整理,丰富了加权弱随机森林模型的特征集,有利于提高算法的分类精度。

4 地物点分类

4.1 随机森林原理

随机森林算法在 2001 年由 Breiman^[16] 提出,是一种基于 Bagging 算法的特征聚合算法。该算法是一种基于传统决策树分类器的集成学习模型,它通过 bootstrap 抽样方法随机生成多个子训练集,并通过随机抽样方法生成特征集;接着利用基尼系数对特征进行量化并构建决策树训练;最后输入测试样本,对所有决策树得到的类别进行归属并投票,得到预测结果。

4.2 算法改进

4.2.1 理论依据

随机森林是由一系列决策树构成。其边缘函数 $P_{\text{mar}}(\mathbf{Q}, \mathbf{V}_T)$ 可以表示为

$$P_{\text{mar}}(\mathbf{Q}, \mathbf{V}_T) = \text{ave}_k o(\mathbf{V}_T) - \max_{\mathbf{V}_F \neq \mathbf{V}_T} \text{ave}_k o(\mathbf{V}_F), \quad (6)$$

式中: \mathbf{Q} 为输入特征向量; o 为结果评价函数; \mathbf{V}_T 为正确分类向量; \mathbf{V}_F 为错误分类向量; ave_k 为均值函数。它的泛化误差是其边缘函数小于 0 的概率,表示为

$$E \leq \frac{\bar{\rho}(1-s^2)}{s^2}, \quad (7)$$

式中: ρ 为决策树的相关系数; s 为分类精度。由(7)式可知,随机森林的泛化误差与决策树之间的相关性有关。因此,通过相关系数对决策树进行筛选,得到相关性小、精度高的决策树组成新的随机森林模型,从而提高分类性能。

4.2.2 降低决策树之间相关性

首先构建初始随机森林模型。本文通过对样本集 M 进行 m 次有放回采样得到训练子数据集,对特征集 N 进行 n 次不放回采样,得到 n 个候选特征作为特征集;对 m 个样本、 n 个特征构建 CART 树,重复 l 次,得到 l 棵决策树 T ,组成初始随机森林模型。

1) 构建最大互信息系数矩阵

假设特征 f_1, f_2 将点云空间划分为四个部分,如图 2 所示。用落在每个空间点的频率作为概率的估计值计算 $p(x, y)$ 、 $p(x)$ 和 $p(y)$,接着通过下式计算两个特征的互信息量。

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_{10} \frac{p(x, y)}{p(x)p(y)}, \quad (8)$$

式中: $p(x)$ 和 $p(y)$ 分别是随机变量 X 和 Y 的边缘概率; $p(x, y)$ 是随机变量 X 和 Y 的联合概率。由

于有多种网格化方案(图 2 中有 a、b、c 三种方案),因此可以找到一种方案使互信息量达到最大。接着计算得到两个特征的最大互信息系数。

$$\text{MIC}(x, y) = \max_{a \neq b} \frac{I(x, y)}{\log_2 \min(r, c)}, \quad (9)$$

式中: $I(X, Y)$ 为特征之间的互信息量; r 和 c 分别表示 f_1 和 f_2 方向上划分的网格数目。

最后对样本所有的特征进行迭代,两两计算得到最大互信息系数矩阵。通过最大互信息系数矩阵,我们可以很方便判断两个特征的相关性。最大互信息系数越大,两个特征越相关。

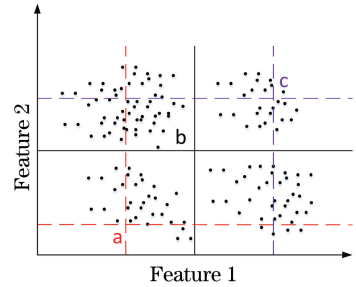


图 2 网格化示意图

Fig. 2 Feature grid diagram

2) 求决策树相关系数

构成决策树的基本单元:特征集和输入样本。因此,在计算决策树相关系数时,可分解为两部分:特征集相关系数和样本相关系数。首先,计算特征集相关系数 C_f 。由于决策树有 n 个特征,通过特征最大互信息系数矩阵计算对应节点特征的最大互信息系数,并求均值作为决策树特征集的相关系数 C_f 。

其次,计算输入样本相关系数 C_s 。假设决策树 T_i, T_j 的训练数据集为 m_i, m_j ,输入样本的相关性主要从样本之间的重合度进行考虑,也就是训练子数据集的重合度。将两个数据集不同点的总个数与训练子数据集样本数的比值作为输入样本的相关系数 C_s ,表示为

$$C_s = \frac{\cup(m_i \neq m_j)}{m}, \quad (10)$$

式中: $\cup(m_i \neq m_j)$ 表示 i, j 不同的样本点数; m 表示训练子数据集样本数。

最后,计算决策树相关系数 C_{XY} 。决策树之间的相关系数值可以通过将输入样本的相关系数值 C_s 与特征的相关系数 C_f 进行求和,得到决策树的相关系数 C_{XY} 。 C_{XY} 的值越大,两棵决策树之间的关联性越大。

3) 构建弱相关随机森林模型

首先计算所有决策树之间的相关系数,其次将相关系数大于阈值 δ 的决策树提取出来构成决策树集,最后选择每个集合中误差最小的决策树构成弱相关随机森林模型。计算公式为

$$E_{err} = \sum_{i=1}^k [h(x_i) \neq t(x_i)]/k, i = 1, 2, \dots, L, \quad (11)$$

式中: $h(x_i)$ 为决策树对节点 x_i 的分类结果; $t(x_i)$ 为节点 x_i 的真实分类结果; k 为子训练数据集的总节点数。

4.2.3 对各决策树结果加权投票

传统的随机森林算法在进行投票时,会对各决策树的结果进行等权重统计。本文通过计算各个决策树的分类精度,对决策树的结果进行加权,提高算法的准确度。由于随机森林算法有 36.8% 的 OOB (out-of-bag) 袋外数据,因此我们使用 OOB 数据作为验证集,对训练得到的 l 棵决策树进行评估,得出

精度向量: $V_{acc} = [a_{acc1} \ a_{acc2} \ \dots \ a_{accL}]$ 。

通过对 OOB 数据中的每一点进行分类,最后得到每个决策树的精度。根据精度向量得到权重向量 $W = [\omega_1 \ \omega_2 \ \dots \ \omega_L]$ 。 ω_i 的计算方式为

$$\omega_i = a_{acc_i}, \quad (12)$$

式中: ω_i 为决策树 i 的权重; a_{acc_i} 为决策树 i 的分类精度。

4.3 算法描述

加权弱相关随机森林模型如图 3 所示。从图中可以看出,算法首先将原始数据分为三类:训练数据、验证数据和测试数据。接着,对特征和训练数据进行随机采样,构建初始随机森林模型,利用验证数据集得到各个决策树分类精度。然后,计算特征之间的相关性,得到特征最大互信息系数矩阵,将相关系数大于阈值的决策树提取出来构成决策树集,选择精度最高的决策树构建新的随机森林。最后,通过加权投票算法对各个决策树分类结果加权投票,得到加权弱相关随机森林模型。

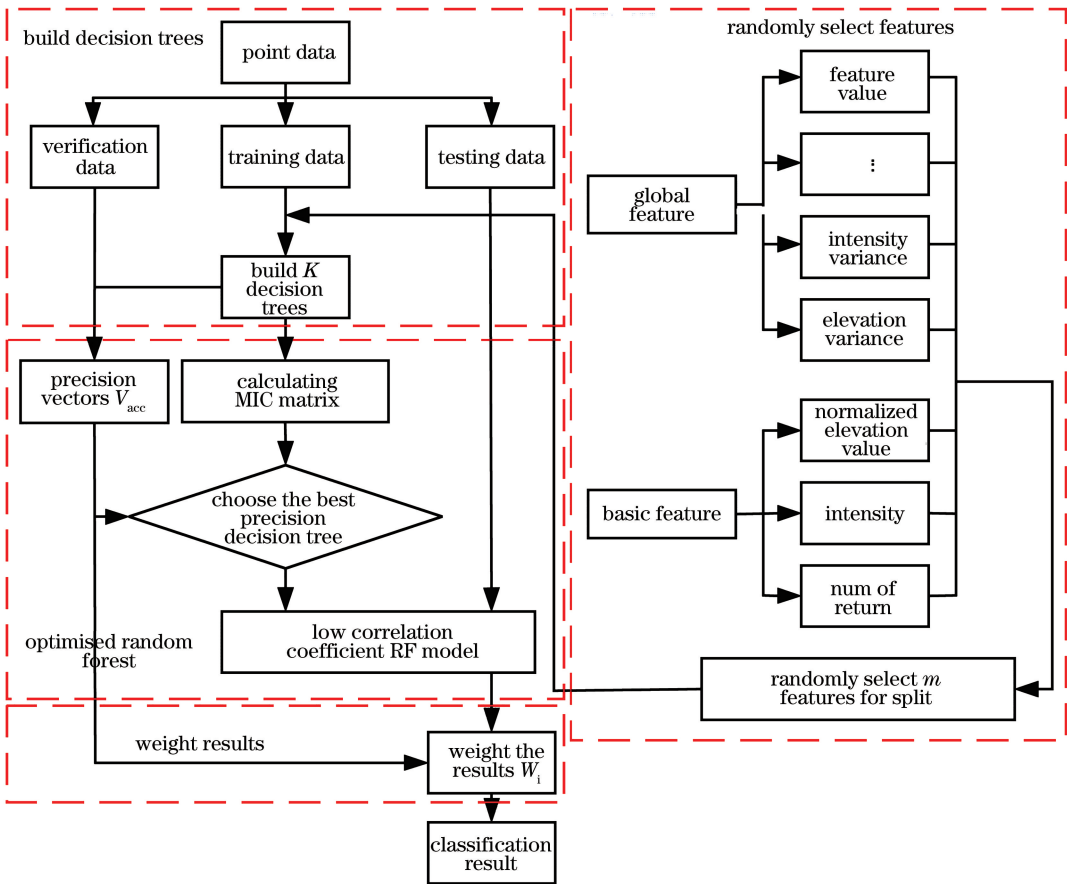


图 3 加权弱相关随机森林模型

Fig. 3 Weighted weakly correlated random forest model

5 实验与分析

5.1 数据集简介

本文使用国际摄影测量与遥感学会 (ISPRS) 提供的 Vaihingen 3D 语义数据集, 主要用于对新方法进行测试和评估^[2]。数据集包括机载高光谱图像, LiDAR 点云数据, 数字表面模型 (DSM)。整个数据

集是由 Leica ALS50 系统获得的, 平均点密度为 4 point/m², 平均点距为 0.66 m, 包含了回波次数和回波强度等数据。

5.2 模型评价指标

为了高效评估算法的准确性, 本文引入了单类分类精度 A_{CA} 、总体分类精度 A_{OA} 、kappa 系数 K 三个指标进行分析。计算公式为

$$A_{CA} = \frac{T_p^{(i)}}{N_p^{(i)}}, \quad (13)$$

$$A_{OA} = \sum_{i=1}^n \frac{T_p^{(i)}}{N_p}, \quad (14)$$

$$K = \frac{N_p \sum_{i=1}^m (T_p^{(i)}) - \sum_{i=1}^m [(T_p^{(i)} + F_p^{(i)}) \times (T_p^{(i)} + F_p^{(i)})]}{N_p^2 - \sum_{i=1}^m [(T_p^{(i)} + F_p^{(i)}) \times (T_p^{(i)} + F_p^{(i)})]}, \quad (15)$$

式中: $T_p^{(i)}$ 表示第 i 类分类正确的个数; $F_p^{(i)}$ 表示第 i 类分类错误的个数; $N_p^{(i)}$ 表示第 i 类点云的总点数; N_p 表示点云总点数; $i=1, 2, \dots, n$, n 表示总共有 n 个种类。单类分类精度 A_{CA} 用于衡量某一类物体的分类情况; 总体分类精度 A_{OA} 和 kappa 系数 K 用于衡量整个算法的分类情况。

5.3 实验结果

首先对点云数据进行滤波处理。图 4 为点云数据高程分布图, 可以看出, 通过归一化处理, 地理环境对点云的影响大大降低, 建筑物的高程特征分布较原始点云数据更为明显。因此, 对点云数据归一化处理可以提高模型的分类效果。接着, 对改进布

料算法的参数进行设置, 一般情况下, 刚性系数的取值范围为 1~3, 当地面比较平坦时, 取值为 3, 此处设置刚性系数为 3。经过测试的实验数据集在迭代次数为 400~500 之间时达到收敛, 故实验设置迭代次数 $N=500$ 。滤波效果如图 5 所示, 图 5(a1) 和 (a2) 为使用传统布料滤波算法对点云滤波的结果, 图 5(b1) 和 (b2) 为使用改进布料滤波算法滤波结果。从图中可以看出, 改进布料滤波算法通过对点云归一化来消除地形影响, 并采用自适应点云密度方法生成布料网格, 从而对建筑物周边的点云以及孤立地面点进行滤波, 有更好的效果, 相比传统布料算法分类精度有很大的提高。

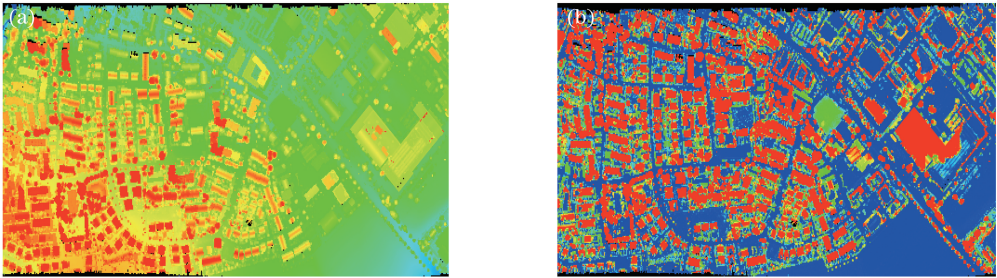


图 4 点云数据高程分布。(a)未归一化点云数据;(b)归一化点云数据

Fig. 4 Point clouds elevation distribution. (a) Unnormalized point cloud data; (b) normalized point cloud data

为了评估布料滤波算法对分类结果的影响, 实验用改进布料滤波算法和弱相关随机森林模型对点云进行滤波和分类, 同时对未进行布料滤波的点云数据直接使用改进弱相关随机森林模型进行分类。对比两种情况下的分类精度, 具体结果如表 1 所示。

通过分析可知, 直接利用改进弱相关随机森林模型的总体分类精度为 0.88, kappa 系数为 0.85; 而用布料滤波算法和弱相关随机森林模型的总体分类精度为 0.91, kappa 系数为 0.90, 说明采用阶层式分类方法, 即使用改进布料滤波算法可以有效的提高

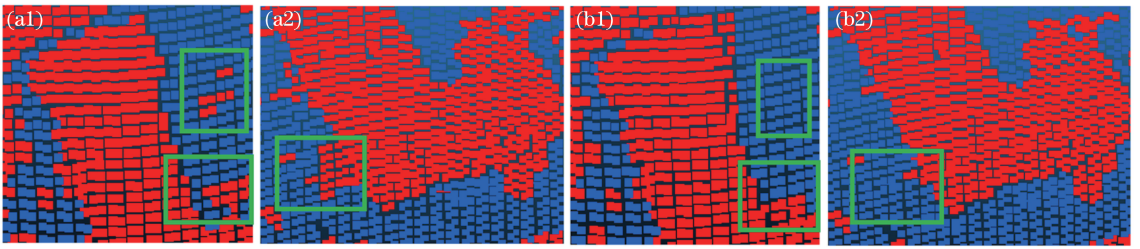


图5 地面点分类结果。(a)传统布料滤波算法;(b)改进布料滤波算法

Fig. 5 Classification results of ground points. (a) Traditional CSF algorithm; (b) improved CSF algorithm

分类精度。从表中可以看出,人造物的单类分类精度由 0.81 提高到 0.87,地面点的单类精度由 0.94 提高到 0.96,说明由于人造物点与地面点特征较接

近,如果直接利用改进弱相关随机森林模型进行分类,则会造成误分类;而通过布料滤波算法过滤地面点后,可以有效提高分类精度。

表1 两种模型实验指标对比

Table 1 Comparison of experimental indicators of the two models

Item	Improved RF					OA	K
	Ground	Building	Vegetation	Artificial	Artificial		
Ground	0.94	0.00	0.00	0.06			
Building	0.00	0.89	0.11	0.00		0.88	0.85
Vegetation	0.00	0.11	0.86	0.03			
Artificial	0.07	0.03	0.09	0.81			
Item	ICSF+Improved RF					OA	K
	Ground	Building	Vegetation	Artificial	Artificial		
Ground	0.96	0.00	0.00	0.04			
Building	0.00	0.91	0.09	0.00		0.91	0.90
Vegetation	0.00	0.08	0.9	0.02			
Artificial	0.00	0.03	0.10	0.87			

5.4 算法对比分析

为了对算法性能进一步评估,将本文算法与传统的随机森林算法、SVM 分类器进行比较,结果如表 2 所示,具体分类结果见图 6,图 6(a)~(d)分别为原始分类结果、传统 RF 算法分类结果、SVM 算法分类结果以及本文算法分类结果。从表 2 可以看出,本文通过改进布料滤波算法和弱相关随机森林模型对点云进行阶层式分类,地面点的分类精度为 0.96,高于传统的随机森林算法,传统的随机森林算法更容易将人造物判别为地面点。本文算法使植被的分类精度有一定的提升,与传统随机森林算法和 SVM 算法相比,分类精度由 0.83 和 0.88 提高到了 0.91;充分说明了通过降低决策树之间的相关性可

以有效提高植被的分类精度。在人造物分类方面,本文算法的分类精度为 0.87,也优于 SVM 算法。从总体分类精度来看,本文算法可以达到 0.91,高于传统随机森林算法的 0.86 和 SVM 算法的 0.89。本文算法的 kappa 系数为 0.90,耗时小于前两种传统算法。传统的随机森林算法由于训练数据集较大,决策树数目较多,因此运行时间较长;SVM 算法在样本过大时运行速度较慢,主要是由于在二次规划求支持向量时需要进行矩阵运算,当样本个数过多时,矩阵运算消耗的内存以及时间随之增加,导致模型建模时间较长。因此,本文算法在运行效率上相对于其他算法都有所提高,并且泛化能力较强,有效提高了分类精度和算法效率。

表2 不同特征组合的分类精度

Table 2 Classification accuracy of different feature combinations

Model	CA				OA	K	Time /s
	Ground	Building	Vegetation	Artificial			
Traditional RF	0.91	0.86	0.83	0.81	0.86	0.82	27.75
SVM	0.94	0.89	0.88	0.82	0.89	0.87	28.20
ICSF + Improved RF	0.96	0.90	0.91	0.87	0.91	0.90	23.56

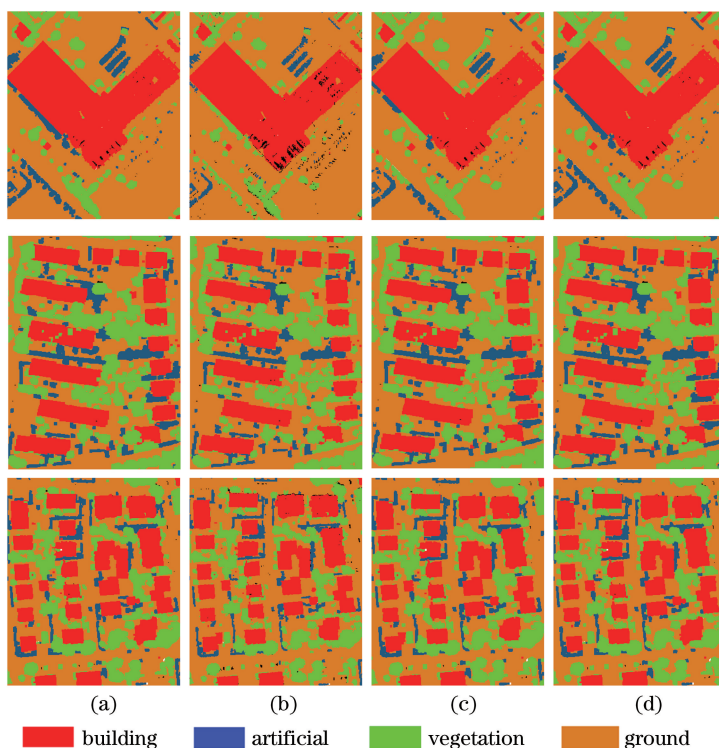


图6 不同算法分类结果。(a)原始分类结果;(b)传统 RF 算法分类结果;(c) SVM 算法分类结果;(d)本文算法分类结果

Fig. 6 Classification results of different algorithms. (a) Correct classification results; (b) classification results of traditional RF algorithm; (c) classification results of SVM algorithm; (d) classification results of proposed algorithm

6 结 论

本文针对点云数据精确分类的问题,提出一种基于 ICSF 和加权弱相关随机森林的点云分类算法。实验结果表明,通过对地面进行归一化处理以及自适应布料网格化后,ICSF 算法可以有效且快速的对点云进行精确滤波;通过计算决策树之间的 MIC 矩阵进而得到相关系数,选择精度较高的决策树,降低对决策树之间相关性可以有效降低随机森林的泛化误差;通过加权使精度高的决策树占的比重更大;与传统算法相比,精确率和效率都得到了提升,解决了传统分类算法分类精度不高的问题。下一步工作可以将点云数据与高光谱数据结合,进一步提高模型的准确率。

参 考 文 献

[1] Zeng Q H, Mao J H, Li X H, et al. Application of the hierarchy classification to point cloud of airborne LIDAR [J]. Science of Surveying and Mapping, 2008, 33(1): 103-105, 249.

曾齐红, 毛建华, 李先华, 等. 机载激光雷达点云的阶层式分类[J]. 测绘科学, 2008, 33(1): 103-105, 249.

- [2] Niemeyer J, Mallet C, Rottensteiner F, et al. Conditional random fields for the classification of lidar point clouds [J]. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012, XXXVIII-4/W19: 209-214.
- [3] Niemeyer J, Rottensteiner F, Soergel U. Classification of urban LiDAR data using conditional random field and random forests [C]//Joint Urban Remote Sensing Event 2013. April 21-23, 2013, Sao Paulo, Brazil. New York: IEEE Press, 2013: 139-142.
- [4] Vosselman G. Slope based filtering of laser altimetry data[J]. Remote Sensing, 2000, 33: 935-942.
- [5] Vosselman G, Gorte B G H, Sithole G, et al. Recognising structure in laser scanner point clouds [M]. Thies M, Koch B, Spiecker H, et al, ISPRS 2004 : proceedings of the ISPRS working group VIII/2 : laser scanning for forest and landscape assessment, Freiburg, October 3-6, 2004. Freiburg: University of Freiburg, 2004: 33-38.
- [6] Kilian J, Haala N, Englich M. Capture and evaluation of airborne laser scanner data [J]. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 1996, 31:

- 383-388.
- [7] Sithole G, Vosselman G. Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2004, 59(1/2): 85-101.
- [8] Zhang W M, Qi J B, Wan P, et al. An easy-to-use airborne LiDAR data filtering method based on cloth simulation[J]. *Remote Sensing*, 2016, 8(6): 501.
- [9] Meng X L, Wang L, Currit N. Morphology-based building detection from airborne lidar data [J]. *Photogrammetric Engineering & Remote Sensing*, 2009, 75(4): 437-442.
- [10] Mallet C, Bretar F, Roux M, et al. Relevance assessment of full-waveform lidar data for urban area classification[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011, 66(6): S71-S84.
- [11] Guo L, Chahata N, Mallet C, et al. Relevance of airborne lidar and multispectral image data for urban scene classification using random forests[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011, 66(1): 56-66.
- [12] Coussement K, van den Poel D. Churn prediction in subscription services: an application of support vector machines while comparing two parameter-selection techniques [J]. *Expert Systems with Applications*, 2008, 34(1): 313-327.
- [13] Battiti R. Using mutual information for selecting features in supervised neural net learning [J]. *IEEE Transactions on Neural Networks*, 1994, 5(4): 537-550.
- [14] Amiri F, Rezaei Yousefi M, Lucas C, et al. Mutual information-based feature selection for intrusion detection systems [J]. *Journal of Network and Computer Applications*, 2011, 34(4): 1184-1199.
- [15] Zhang R, Li G Y, Li M L, et al. Classification of LiDAR point clouds based on PCA-BP algorithm[J]. *Bulletin of Surveying and Mapping*, 2014(7): 23-26. 张蕊, 李广云, 李明磊, 等. 利用 PCA-BP 算法进行激光点云分类方法研究 [J]. *测绘通报*, 2014(7): 23-26.
- [16] Breiman L. Random forests [J]. *Machine Learning*, 2001, 45(1): 5-32.
- [17] Mohd Salleh M R, Abd Rahman M Z, Ismail Z, et al. Revised progressive morphological method for ground point classification of airborne lidar data [J]. *International Journal of Built Environment and Sustainability*, 2019, 6(1/2): 31-38.
- [18] Cao Q, Zhong Y F, Ma A L, et al. Urban land use/land cover classification based on feature fusion fusing hyperspectral image and lidar data [C]//IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. July 22-27, 2018, Valencia, Spain. New York: IEEE Press, 2018: 8869-8872.
- [19] García-Sopo M Á, Cuartero A, Rodríguez P G, et al. Hyperspectral and lidar data integration and classification [C]//2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). July 26-31, 2015, Milan, Italy. New York: IEEE Press, 2015: 57-60.