A two-stage ALS point cloud segmentation framework for urban areas

Wei Han¹, Daqing Huang¹,²,* and Cheng Xu¹,²

¹ College of Electronic and Information Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing, China
² Key Laboratory of Advanced Technology for Medium and Small UAV, Ministry of Industry and Information Technology, Nanjing, China

*Corresponding author e-mail: mxyacc@nuaa.edu.cn

Abstract. In this paper, a novel methodology for semantic classification of airborne laser scanning point clouds is introduced in this paper. A two-stage framework integrating point-based classification and cluster-based rules is proposed with a general end to end processing scheme, which is from scattered airborne laser scanning (ALS) data to points with semantic labels. In the first stage, energy function consists of a point-wise-based soft labelings term obtained by random forest classifier and a spatially smooth labelling term. Then, a constrained mean-shift-based clustering algorithm combined with semantic rules is proposed to refine the hard labelings of class-wise point clouds, which can be as a post-processing stage. To verify the effectiveness of our method, the classification results derived for an ALS benchmark dataset are shown in Section 3. Base on the semantic classification results, the two-stage framework can be demonstrated with high accuracy and high plausibility in comparison to some classic approaches.

1. Introduction

The interpretation ability for the local environment, which can be acquired from a sensor platform, represents a key prerequisite for some high-level tasks, such as automatic three-dimensional (3D) city modeling, self-navigation based on the scene or detection task for specific object. In particular, the interpretation benefits from the semantic labeled point clouds. Then, in this paper, the semantic classification of ALS point cloud is exploited for some high-level tasks.

Airborne LiDAR (Light Detection and Ranging) is a laser scanning device and can be used to collect the 3D coordinates of the survey area from laser range measurements, which is a particularly useful technology. Due to the fact that airborne laser scanning (ALS) datasets are significantly more noises, uneven and unorganized, the accuracy of different types of objects segmentation is far from being desired and it is a difficult task to extract them reliably. In this paper, we focus on the semantic classification of irregularly distributed ALS point cloud, assigning a semantic label for each point, without image data providing spectral information.

In lately research, the semantic classification technology for point clouds has been applied in a quantity of investigations [1]-[4]. It can be classified three types: point-based group, segment-based group, and point-image-based group which is according their essential processing units’ category. A synopsis especially the specific applications are provided below. The CRF framework can be inserted a random forest (RF) detector which based on probabilities computed through unary and pairwise
potentials of CRF in [5]. It can be extracted more discriminative features with mixing airborne LiDAR and images that combines geometric features and radiometric features by [6]. [7] used a standard supervised classification scheme as input to simplify a various of specific geometric low-level features during all processing points with an optimized neighborhood method. [8] and [9] proposed a multi-stage inference procedure which including point cloud statistics and relational information by various scales. A regularization to capture a smooth labeling by [10] which can be interacted with points themselves in nearby from the training data. All the above methods have different kinds of constraints on the semantic classification in irregularity and sparsity ALS data cases.

Based on the experiments results, the brief and efficient approach proposed in this work has been shown a good performance on ALS point cloud, which is demonstrated and evaluated on a benchmark data set with complex urban scenes. During the semantic classification, five classes are discerned and a quantitative evaluation is carried out with ground truth.

This paper is organized as follows: Section 2 presents our methodology with a brief description of energy function formalization and post-processing stage. Then, Section 3 contains the evaluation for the ALS point cloud semantic classification and the paper concludes in Section 4.

2. Methodology

The goal of point cloud semantic classification is to assign a semantic label \( l_L \in L \) to each point \( i \). \( V \) and \( L \) can be denoted as the finite set of ALS points to label and the label set which contains unique semantic labels for point cloud respectively.

2.1. Energy Function Formalization

In the first stage, the probability that a label \( l \) is assigned to the point \( i \) is defined with a RF classifier, in which cross-validation is utilized for parameters decision [11]. Then, to measure the similarity between two different distributions, Kullback-Leibler divergence \( KL(P, Q) \) is employed. On the other hand, spatially smooth labeling term can be represented by the total variation, which is similar with Potts model. The soft labelings term and spatially smooth term are integrated into energy function formalization, as follows,

\[
Q \in \arg \min_{Q \in \Omega} \{ \sum_{i \in V} E(P_i, Q_i) + \sum_{(i,j) \in E} E(Q_i - Q_j) \}
\]

where \( P \) is a global soft labeling distribution, \( Q \) the distribution of an optimization result and \( \Omega \) the search space of \( Q \).

2.1.1. Soft Labelings Term. A probability is assigned to each label in a soft labelling results, and consequently more information can be contained. Our goal is to obtain a hard labelling, i.e. each point with a specific label, which is also the main objective in semantic classification. In this step, the soft labelings term approximation minimizes the KL divergence

\[
D(P \| Q) = \sum_k P_k (\log(1 / Q_k) - \log(1 / P_k))
\]

We focus on optimizing this term in regard to \( Q \), so we can reserve the part with \( Q \) and neglect the constant component. Then, a convex combination can be derived with a uniform parameter \( \gamma \in [0,1] \).

\[
E(P, Q) = \sum_{l \in L} \left( \frac{\gamma}{\text{size}(L)} + (1 - \gamma) P_{(i,l)} \right) \cdot \log \left( \frac{\gamma}{\text{size}(L)} + (1 - \gamma) Q_{(i,l)} \right)
\]

where \((i,l)\) represents the \( i \)th point with label \( l \).
2.1.2. Spatially Smooth Labelling Term. This term favours solutions which are spatially smooth, i.e. same labels are mostly assigned to the adjacent nodes. For the soft search space, $\ell_0$-norm-based total variation is a convex relaxation of Potts model, which can be formulated by the cumulative sum of the absolute differences between adjacent distributions. The spatially smooth labelling term can be described as:

$$E(Q_i, Q_j) = \sum_{i \in L} |Q_{(i,j)} - Q_{(j,i)}|$$  \hspace{1cm} (4)$$

2.2. Post-Processing

In the first stage, a hard labeling result of ALS data can be obtained. Though the overall accuracy of the semantic segmentation result has been improved, some information on classes is lost, especially small size classes. In order to reserve more semantic information and further improve the overall accuracy, a post-processing stage is proposed.

For dealing with ALS data, which general be a large-scale scene, Cloth Simulation Filter (CSF) is used to remove ground points. We focus on the non-ground points in this stage.

2.2.1. Constrained Mean-Shift-based Clustering. Firstly, an over-segmentation for ALS point clouds is performed with the mean-shift algorithm, in which kernel density estimation is employed to locate the maximum values in a density function. In order to improve the efficiency, a down-sampling step is performed on the non-ground points. After clustering, cluster set $C$ can be obtained. Then two constraints are defined to merge clusters into a high level, i.e. one cluster may represent an object.

- **Constraint 1: Local Connectivity**
  $$d(p_i, p_j) \leq th_1 \quad \text{s.t.} \quad (p_i, p_j) = \arg \min_{(p_1, p_2)} d(p_1, p_2) : p_1 \in c_i, p_2 \in c_j$$  \hspace{1cm} (5)$$
  
  where $d(\cdot)$ is the Euclidean distance of two points, $c_i$ and $c_j$ clusters in the set $C$, $th_1$ the threshold of the constraint.

- **Constraint 1: Direction Correlation**
  $$\|\mathbf{N}(c_i) \cdot \mathbf{N}(c_j)\| \leq th_2$$  \hspace{1cm} (6)$$

  where $\mathbf{N}(c_i)$ and $\mathbf{N}(c_j)$ are normal vectors estimated from the covariance matrix of $c_i$ and $c_j$ clusters, $th_2$ the threshold of the constraint.

2.2.2. Semantic Rules. Generally, natural and artificial objects have been closely related to different semantic information, which can be verified by semantic rules. To define the semantic rules, cluster-based features are extracted. Based on these features, the useful rules of each class for our approach are shown as follows.

- **Buildings** are characterized by a certain range of height and a relatively stable angle between the vertical direction and the normal vector of clusters. The roughness of building is the lowest among all the classes except ground points.

- **A car cluster** should have a small size of point. For a query car cluster, the distribution of ground points is measured by a cylindrical region surrounding the cluster, which can distinguish between small and large objects.

- **The size of cluster and the distribution of ground points in a cylindrical region can also be used to prevent the misjudgement between car and low vegetation, which have similar heights and roughness. Height above the ground is important information for the high vegetation, e.g. trees. With roughness, the building and high vegetation can be well classified.**

  Based on the above semantic rules, the misjudged point labels are able to be corrected, which can improve the overall accuracy and obtain a reliable semantic segmentation result.
3. Experimental Results

In our experiments, the labeled benchmark ALS dataset *GML Dataset* provided by Graphics and Media Lab [12] is employed to evaluate the performance of the proposed method. The *GML Dataset*, acquired from an ALTM 2050 system, is publicly available without DTM information and probably 2.077M labeled points are contained in it. There are five semantic classes, i.e. *Ground, Building, Car, Tree* and *Low Vegetation*, in this dataset, as indicated in Table 1.

In this experiment, we first extracted 26 point-based features calculated with optimal neighbors. With the permutation importance measurement, the feature importance is derived, which can be shown in Fig.1. Based on that, meaningful features can be selected for the RF classifier. The following features are used:

1) tensor-based features: linearity( $L$ ), scattering( $S$ ), planarity( $P$ ), Omni-variance( $O$ ), anisotropy( $A$ ), Eigen-entropy( $E$ ), change of curvature( $CC$ ), sum of eigenvalues($\sum$), sum and ratio of 2d-based eigenvalues($\sum_{2d}$, $Ra$);

2) geometric-based features: height above ground( $H$ ), maximum difference and variance of height among optimal neighbors( $\Delta H$, var$_H$ ), curvature ( $C$ ), variance of curvature( var$_C$ ), roughness( $R$ ), variance of roughness( var$_R$ ), density and 2d-based density($D$, $D_{2d}$ ), variance of density and 2d-based density( var$_D$, var$_D_{2d}$ ), radius and 2d-based radius( $r$, $r_{2d}$ ), normal vector relative( $N$ ), variance of normal vector relative( var$_N$ ).

| Class        | Training Data | Test Data |
|--------------|---------------|-----------|
| Ground       | 557,142       | 439,989   |
| Buildings    | 98,244        | 19,592    |
| Car          | 1,833         | 3,235     |
| Tree         | 381,677       | 531,852   |
| Low vegetation | 35093       | 7758      |

The RF-based classified result is shown as Fig.2(b), while Fig.2(a) is the ground truth of *GML test datasets*. For simplicity, a portion classification result of *GML test datasets* is presented, as shown in Fig.2. Fig.2(c) shows the result of the first stage, which is to minimize the energy function described in Sec. (2.1) with a-expansion algorithm. From Fig.2(c), we can see that the semantic segmentation result is smoother than result shown in Fig.2(b). However, some semantic information are lost, which lead to an under-segment result. In the second stage, constrained Mean-Shift-based clustering and semantic rules are performed to revise some misjudged labels in first stage. The cluster result of non-ground points is shown in Fig.3. Combined the hard labeling obtained from the first stage with semantic rules defined in the second stage, the final semantic segmentation result can be shown in Fig.2(d), which shows a semantic classification result with high plausibility.

For the evaluation, five commonly used measures (in %), i.e. overall accuracy( $OA$ ), precision( $P$ ), recall ( $R$ ), Kappa coefficient( $Ka$ ) and F1-score, are employed, which can be derived based on the confusion matrix. Different methods are employed to test our proposed method and the evaluation results are provided in Tab.2. In general, the overall performance based on the two-stage method, in which the influence of clutter points can be reduced and more semantic information can be exploited, is quite approving on the ALS point cloud semantic classification.
Figure 2. Semantic segmentation results.

Figure 3. Constrained Mean-Shift-based clustering in the second stage.

Table 2. Class-wise overall accuracy (OA), precision (P), Kappa coefficient (Ka), recall (R), and F1-score for the GML dataset with different methods.

| Methods             | OA  | P   | Ka  | R   | F1-score |
|---------------------|-----|-----|-----|-----|----------|
| RF-based            | 81.6| 51.6| 70.7| 64.3| 52.6     |
| CRF-based           | 89.4| 51.3| 80.9| 57.3| 65.7     |
| LBP-based           | 86.0| 57.6| 77.1| 67.7| 57.3     |
| a-expansion-based   | 87.8| 58.7| 79.7| 67.0| 57.6     |
| Our method          | 91.5| 69.1| 80.9| 77.6| 66.7     |

4. Conclusions

In this paper, a two-stage framework for ALS data semantic classification is presented. The novelty integration of point-wise soft labeling optimization, constrained clustering and semantic rules for the irregularly distributed ALS data, shows a approving semantic classification result. Firstly, the soft labeling result is derived from a RF classifier for a considered ALS point cloud. Then, a hard labeling based on the energy function established by soft labeling term and spatially smooth labeling term can be obtained as an initial labeling. The constrained mean-shift-based clustering allows to derive cluster-wise geometric information, which can be utilized for the definition of semantic rules. Some failure cases shown in a variety of papers can be addressed to further improve the derived semantic classification results. However, a limit is existed in our approach, i.e., thresholds defined in this paper have different values for different kinds of ALS data.
Future work will be focused on integrating the mixed semantic information into single formulation. Furthermore, we will try to improve the final classification with deep learning strategy. Then, reconstruction of objects can be further investigated based on the semantically enriched ALS point clouds.

5. Acknowledgments

This work was supported by Funding of Jiangsu Innovation Program for Graduate Education, the Fundamental Research Funds for the Central Universities (Grant No.KYLY15_0287); National Natural Science Foundation of China (61601222); National Key Research and Development Project (Grant No.2017YFC0822404); China Postdoctoral Science Foundation (2018M632303) and Natural Science Foundation of Jiangsu Province (BK20160789).

References

[1] Nguyen A, Le B. 3D point cloud segmentation: A survey[C]// The 6th International Conference on Robotics, Automation and Mechatronics (RAM). IEEE, 2013
[2] Shujuan Y, Keshu Z and Yongshe S. Classification of Airborne LiDAR Point Cloud Data Based on Multiscale Adaptive Features[J]. Acta Optica Sinica, 2019, 39(2):0228001
[3] Weinmann M, Hinz S, Weinmann M. A Hybrid Semantic Point Cloud Classification-Segmentation Framework Based on Geometric Features and Semantic Rules[J]. Pfg – Journal of Photogrammetry Remote Sensing & Geoinformation Science, 2017, 85(3):183-194
[4] Jie S, Zulong L. Airborne LiDAR Feature Selection for Urban Classification Using Random Forests[J]. Geomatics and Information Science of Wuhan University, 2014, 38(11):207-212
[5] Niemeyer J, Rottensteiner F, Soergel U. Contextual classification of lidar data and building object detection in urban areas[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 87:152-165
[6] Lodha, S.K., Fitzpatrick, D.M., Helmbold, D.P., Aerial LiDAR data classification using AdaBoost[C]. In: Proc. 6th International Conference on 3-D Digital Imaging and Modeling (3DIM’07), Montreal, Canada, IEEE, 2007
[7] Weinmann M, Jutzi B, Mallet C. Semantic 3D scene interpretation: A framework combining optimal neighborhood size selection with relevant features[J]. 2014 ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-3
[8] Xiong X, Munoz D, Bagnell JA,HebertM. 3-d scene analysis via sequenced predictions over points and regions[C]. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA’11). IEEE, 2011,2609–2616
[9] Shapovalov R, Vetrov D, Kohli P. Spatial inference machines[C]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’13). IEEE, 2013,2985–2992
[10] Landrieu L, Raguet H, Vallet B, et al. A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 132:102-118
[11] Huan N, Xiangguo L, Jixian Z. Classification of ALS Point Cloud with Improved Point Cloud Segmentation and Random Forests[J]. Remote Sensing, 2017, 9(3):288
[12] Shapovalov R, Velizhev A. Cutting-Plane Training of Non-associative Markov Network for 3D Point Cloud Segmentation[C]// 2011 IEEE International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, IEEE, 2011