Semantic Labeling of Mobile LiDAR Point Clouds via Active Learning and Higher Order MRF

Huan Luo, Cheng Wang, *Senior Member, IEEE,* Chenglu Wen, *Senior Member, IEEE,* Ziyi Chen, Dawei Zai, Yongtao Yu, and Jonathan Li, *Senior Member, IEEE*

***Abstract*— Using mobile Light Detection and Ranging point clouds to accomplish road scene labeling tasks shows promise for a variety of applications. Most existing methods for semantic labeling of point clouds require a huge number of fully supervised point cloud scenes, where each point needs to be manually annotated with a specific category. Manually annotating each point in point cloud scenes is labor intensive and hinders practical usage of those methods. To alleviate such a huge burden of manual annotation, in this paper, we introduce an active learning method that avoids annotating the whole point cloud scenes by iteratively annotating a small portion of unlabeled supervoxels and creating a minimal manually annotated training set. In order to avoid the biased sampling existing in traditional active learning methods, a neighbor-consistency prior is exploited to select the potentially misclassified samples into the training set to improve the accuracy of the statistical model. Furthermore, lots of methods only consider short-range contextual information to conduct semantic labeling tasks, but ignore the long-range contexts among local variables. In this paper, we use a higher order Markov random field model to take into account more contexts for refining the labeling results, despite of lacking fully supervised scenes. Evaluations on three data sets show that our proposed framework achieves a high accuracy in labeling point clouds although only a small portion of labels is provided. Moreover, comparative experiments demonstrate that our proposed framework is superior to traditional sampling methods and exhibits comparable performance to those fully supervised models.**

***Index Terms*— Active learning, conditional random field (CRF), higher order Markov random field (MRF), mobile Light Detection and Ranging (LiDAR) point clouds, semantic labeling.**

I. Introduction

In RECENT years, urban traffic congestions and traffic accidents have increasingly constrained a modern lifestyle and sustainable urban development [1]. To effectively collect road information and gather traffic information for solving those urban transport issues, a large number of sensors, such as infrared sensors, laser sensors, and cameras, are used [2]–[4]. A lot of intelligent applications, including driver assistance and safety warning systems, and autonomous driving, benefit from understanding contextual information about a road and its periphery (e.g., the locations of light poles, trees, and vehicles). Semantic labeling of road scenes, automatically assigning a category label to each basic element (e.g., pixel or point) in road scenes, provides a promising and essential approach to obtain the knowledge about road environments. Over the past few decades, studies on labeling road scenes focused mainly on optical images and videos [5], [6]. The use of optical images and videos to conduct semantic labeling of road scenes is limited, due to illumination conditions, occlusions, distortions, incompleteness, viewpoints, and lack of geospatial information.

With fast-developing Light Detection and Ranging (LiDAR) technologies, large volumes of highly dense and accurate point clouds, which are easily and rapidly acquired by mobile LiDAR systems, provide a new solution to represent road-related information. The collected point clouds exhibit advantages over optical images and videos captured by traditional optical imaging-based systems. By integrating laser scanners with position and orientation systems, mobile LiDAR systems rapidly capture undistorted 3-D point clouds with real-world coordinates of road scenes. Such 3-D point clouds assist in accurate object localization in road scenes. In addition, compared with optical imaging-based systems, mobile LiDAR systems are immune to the impact of illumination conditions. Moreover, with the complementary onboard high-resolution digital cameras, the colorized point clouds provide not only geometric but also texture information essential to image-based semantic labeling. Therefore, in this paper, we focus on semantic labeling of road scenes by using mobile LiDAR point clouds.

To train a statistical model for semantic labeling of point clouds, most existing methods [7]–[11] require a huge

Manuscript received March 2, 2016; revised November 21, 2016, May 13, 2017, and December 30, 2017; accepted February 1, 2018. Date of publication May 2, 2018; date of current version June 22, 2018. This work was supported in part by the Natural Science Foundation of China under Project U1605254 and Project 61771413 and in part by the Collaborative Innovation Center of Haixi Government Affairs Big Data Sharing. (*Corresponding author: Cheng Wang.*)

H. Luo is with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China, and also with the College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China (e-mail: hluo@fzu.edu.cn).

C. Wang, C. Wen, Z. Chen, and D. Zai are with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, Xiamen 361005, China (e-mail: cwang@xmu.edu.cn; clwen@xmu.edu.cn; chenziyicip@163.com; david102812@gmail.com).

Y. Yu is with the Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 223003, China.

J. Li is with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, Xiamen 361005, China, and also with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: junli@xmu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2018.2802935

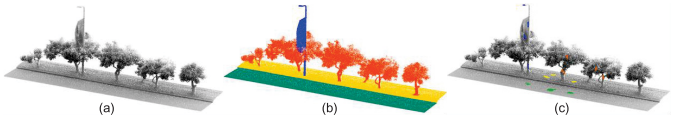


Fig. 1. Example of training data in traditional methods and our proposed method on semantic labeling of point clouds. (a) Unlabeled point cloud scene. (b) Fully supervised training data required by traditional labeling methods. (c) Training data generated by the active learning method. Here, gray represents unlabeled points, and other colors represent manually labeled points.

number of fully supervised complete scenes, in which each 3-D point is manually annotated with a specific category [see Fig. 1(b)]. However, such manual annotations for point clouds are difficult to obtain in terms of cost and time. In addition, it seems impossible to accomplish accurate annotations for each point from a complete scene in some scenarios, e.g., classifying points of overlapping trees and light pole manually [see Fig. 1(a)]. In fact, only a small portion of labeled points from complete scenes determines the parameters of a statistical model. In the machine learning literature, active learning is dedicated to create a minimal training data set from a huge pool of unlabeled data by iteratively selecting valuable samples to query their category labels [12]–[14]. Thus, in this paper, to reduce the cost of manually annotating training data, instead of manually annotating whole point cloud scenes, we present semantic labeling of point clouds by actively and automatically selecting a small portion of unlabeled points for manual annotation [see Fig. 1(c)]. Based on those manually annotated points, a statistical model for semantic labeling of point clouds is learned.

Recently, probabilistic graphical models, e.g., Markov random field (MRF) [15] and conditional random field (CRF) [16], were commonly explored to account for contextual information in semantic labeling of point clouds [8]. Active learning requires frequently retraining a statistical model. Therefore, in our framework, at the model learning stage, due to computational concerns during learning and inference, we choose pairwise CRFs, where unary and pairwise potentials carry category probabilities and contextual information between neighboring variables, respectively.

A lot of work demonstrates that a higher order graphical model, which models long-range interactions between variables, provides more knowledge about the context of a scene and improves the semantic labeling results [10], [11], [18]. Only modeling local interactions among variables by pairwise CRFs is insufficient to encode long-range contextual information among variables and reduces the labeling accuracy. Therefore, in this paper, we propose to use a higher order MRF to refine the labeling results obtained by the pairwise CRFs. However, our active learning method only provides training samples as a set of separated and annotated points rather than fully supervised scenes. Because of lacking fully supervised scenes at training stages, it is challenging to adapt traditional higher order MRFs into label refinement directly. Therefore, in labeling framework, a higher order term not depending on fully supervised training scenes is needed. Inspired by the observation of describing a region with as few categories as necessary, we propose a higher order term named regional label cost term to reduce unnecessary categories by imposing costs on the used categories in labeling a region. The proposed regional label cost term can perform well despite lack of fully supervised training scenes and is suitable to be applied in refining the labeling results inferred by pairwise CRFs learned in active learning procedure.

In this paper, we propose a new framework using active learning and higher order MRF for semantic labeling of mobile LiDAR point clouds. Active learning iteratively selects a portion of unlabeled samples to be manually annotated and creates a minimal training set. Once the creation of training set is finished, a pairwise CRF is learned to classify the unlabeled samples in the road scene of point clouds. To improve the labeling results obtained by a pairwise CRF, we present a higher order MRF, which applies regional label cost terms to explore long-range interactions among variables. Our proposed framework is validated on three data sets of mobile LiDAR point clouds, and the evaluations exhibit the capability of our proposed framework on semantic labeling of point clouds.

The main innovative contributions of this paper to semantic labeling of mobile LiDAR point clouds can be summarized as follows.

1) To avoid annotating the whole training scenes manually and reduce the requirements of manually annotated training samples for labeling point cloud scenes, we introduce active learning to select as few points as possible for manual annotation and to form a minimal training set. To conduct unbiased sampling during active learning procedure, we propose to exploit the neighbor-consistency prior to select the potentially inaccurately labeled samples to be annotated manually.
2) To consider more contextual information into semantic labeling, we propose a higher order MRF method to refine the labeling results obtained by pairwise CRF. The proposed higher order MRF method, which does not require fully supervised training scenes, improves the labeling results by reducing unnecessary categories used in describing a region.

The remainder of this paper is organized as follows. Section II introduces some related work. Section III presents the components of our proposed framework. Section IV reports extensive experimental results and evaluates the performance of the proposed framework. Finally, Section V gives the concluding remarks and hints at plausible future research.

II. Related Work

Most works on semantic labeling of point cloud road scenes focused mainly on exploiting probabilistic graphical models. The pairwise CRF was used to extensively ensure category label consistency between neighboring points [8], [19]–[21]. In [8], a maximum-margin framework is proposed to discriminatively train a pairwise associative Markov networks to annotate the objects of interest. In [20], to reduce redundancy of labeling every individual point, adaptive support regions (supervoxels) are treated as basic units to model a multiscale pairwise CRF. In [21], a patch-based framework was proposed to label road scenes by exploiting object intrinsic properties to transfer category labels from labeled scenes to unlabeled

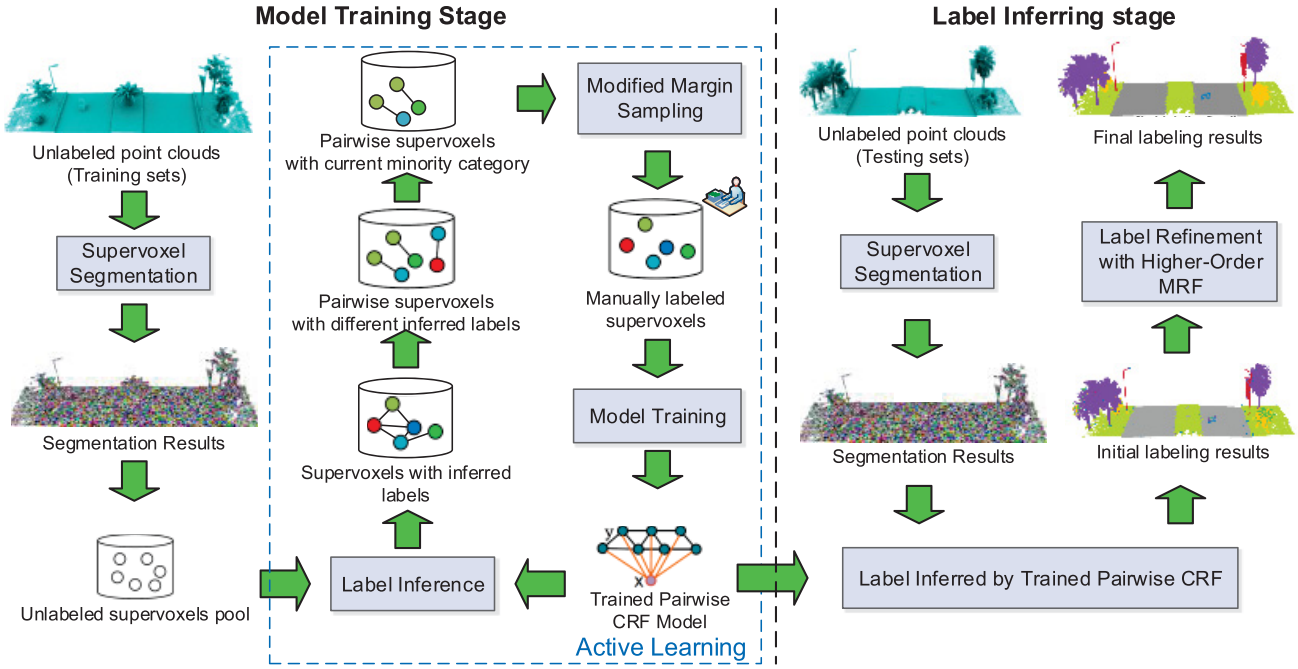


Fig. 2. Overview of our proposed framework for semantic labeling of point clouds. (Different colors represent different categories.)

ones and applying a pairwise CRF model to consider contexts for refining the transferred labels. In [22], random forest (RF) classifiers were learned on the training data automatically generated by exploiting the prior knowledge among classes, and the labeling results were further refined by pairwise CRF. In [23], the weak priors in the street environment were used to conduct automatic generation of training data. Based on those training data, a pairwise CRF-based semantic labeling method was proposed to segment images and scanned point cloud simultaneously. The success achieved by pairwise CRFs notwithstanding, long-range interaction between variables, essential to exploit more contextual information in complex scenarios, is ignored. The Potts model (a higher order graphical model) [24] was used to keep category labels homogeneous in a predefined clique [11]. To allow a portion of inhomogeneous labels in a clique, a robust Potts model [25] was introduced and integrated into the Max-Margin Markov Network (M3N) [10]. In [18], a set of new higher order pattern-based potentials were designed to encode the geometric relationships between different categories within the cliques, rather than simply encourage the nodes in a clique to have consistent labels. Considering a large amount of annotated data required in the past studies, our proposed framework introduces active learning to reduce the large amount of demand on annotated data for the labeling tasks.

Due to the complexity of the probabilistic graphical models, in the semantic labeling area, there were only a few studies [26], [27] on the combination of probabilistic graphical models and active learning strategies. In [26], an expect change strategy was used to find the informative samples, which induce largest expected changes in overall CRF state after revealing their true labels. In margin-based sampling, a loopy belief propagation algorithm [28] was used to exploit both spectral and spatial information to actively select informative samples, where conditional margin of each sample was estimated in a discriminative random field model [27]. Li *et al.* [27] believe that integration of probabilistic graphical models and active learning assists in providing both local and contextual information for selecting informative samples. In our proposed framework, we not only consider the neighboring contexts information to select the most informative samples by using a pairwise CRF model, but also try to keep the diversity of the selected samples to some extent by adding the potentially misclassified samples into the manually annotated training set.

III. Proposed Framework

Section III is organized as follows. An overview of our proposed framework for semantic labeling of mobile LiDAR point clouds is presented in Section III-A. Then, the supervoxel segmentation is described in Section III-B. The active learning is given in Section III-C. Finally, category label refinement with incorporated regional label costs is explained in Section III-D.

A. Overview of the Proposed Framework

Our proposed framework is divided into two stages: model training stage and label inferring stage. As shown in Fig. 2, at the model training stage, unlabeled training point cloud scenes are first oversegmented into spatially consistent supervoxels through the voxel-cloud connectivity segmentation (VCCS) algorithm [29]. After supervoxel extraction, all the unlabeled supervoxels form an unlabeled supervoxels pool. Then, in the pool, active learning is applied to select valuable unlabeled supervoxels to query their category labels.

In addition, those supervoxels, with queried labels, are formed as a training set and used to learn a pairwise CRF model.

At the label inferring stage, initial labeling of unlabeled point cloud scenes is first inferred by applying a trained pairwise CRF. Because long-range interactions in a region cannot be well modeled by using only unary and pairwise potentials in a pairwise CRF model, some mislabeled supervoxels remain in the initial labeling results (see Fig. 2). To refine the initial labeling results, we exploit a higher order MRF model to describe long-range interactions between supervoxels for category label refinement.

B. Supervoxel Segmentation

To reduce the huge computational burden brought by the large amount of points in our data set, supervoxels, instead of the original points, are treated as basic operational units in the proposed framework. The VCCS algorithm is an effective supervoxel generation algorithm [29], where points within each supervoxel have consistent 3-D geometry and appearance. Moreover, supervoxels obtained by the VCCS algorithm can effectively preserve boundary information according to the constraint that each supervoxel cannot flow across the object boundaries. Therefore, it is suitable to directly handle the supervoxel in point cloud labeling tasks. In the proposed framework, given a point cloud scene, we obtain a set of supervoxels using the VCCS algorithm. There are two important parameters: voxel resolution and seed resolution. The voxel resolution is used to define the operable unit of the voxel-cloud space, whereas the seed resolution determines the seed points for constructing initial supervoxels. In this paper, the voxel resolution and seed resolution are set at 0.05 and 0.1 m, respectively.

To describe each supervoxel, we use the following features:

1) Fast Point Feature Histograms (FPFHs) descriptor [30], a rotation-invariant feature, which describes the local surface geometry of points in a supervoxel;
2) spectral features [11] that capture scatter, linearity, and planarity of point distributions in a supervoxel;
3) deviation of the normal vector direction of a supervoxel from the z-axis, which assists in distinguishing between the horizontal and vertical planar surfaces [11];
4) height of the centroid point in a supervoxel;
5) mean RGB color values in a supervoxel.

C. Active Learning

To reduce the manual annotation of training samples, given a pool of unlabeled supervoxels $\mathcal{S}$, active learning iteratively selects a set of unlabeled supervoxels to be manually annotated. The manually annotated supervoxel set, $\mathcal{D}_L \subseteq \mathcal{S}$, is treated as a training set to train a statistical model $\mathbf{w}$. Algorithm 1 gives the main procedure of the active learning algorithm. In Algorithm 1, line 3 trains a statistical model based on current annotated samples $\mathcal{D}_L$. Here, in order to consider contextual information between supervoxels, our proposed framework selects pairwise CRF as a statistical model. Line 4 selects the valuable supervoxel x_s under current CRF model and manually annotates the selected valuable supervoxel. In our proposed framework, we propose a new sampling method called modified margin-based sampling (MMbS) to select valuable supervoxels.

Algorithm 1 Active Learning Algorithm

Input: a pool of unlabeled supervoxels, $\mathcal{S}$
Output: the manually annotated supervoxel set, $\mathcal{D}_L$, and a statistical model, $\mathbf{w}$

1: initialize $\mathcal{D}_L$ by annotating several samples manually
2: **repeat**
3: $\mathbf{w}$ = pairwise_CRF_model_training($\mathcal{D}_L$)
4: x_s = AL_Select_Valuable_Sample($\mathbf{w}, \mathcal{S}$)
5: $\mathcal{S} = \mathcal{S} - x_s$
6: $\mathcal{D}_L = \mathcal{D}_L + x_s$
7: **until** the stopping criterion is met
8: **return** $\mathcal{D}_L$ and $\mathbf{w}$

In the remainder of this section, we first introduce a pairwise CRF model. Second, the proposed sampling method, MMbS, is explained. Finally, the whole procedure of actively selecting valuable samples is described.

1) Pairwise CRF Model: Given a set of supervoxels $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ obtained from point cloud scenes, where N is the number of supervoxels, the semantic labeling tasks predict a labeling, $\mathbf{y} = (y_1, y_2, \ldots, y_N)$, for all the supervoxels $\mathbf{x}$. A category label, $y_i \in \mathcal{L} = \{1, \ldots, L\}$, is assigned to each supervoxel x_i. Here, L is the number of categories.

With these definitions in place, we build the posterior density $p(\mathbf{y}|\mathbf{x})$ of the categories $\mathbf{y}$, given the features of supervoxels, $\mathbf{x}$ by a pairwise CRF model

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x}, \mathbf{w})} \exp(-E_s(\mathbf{x}, \mathbf{y}, \mathbf{w})) \tag{1}$$

where $Z(\mathbf{x}, \mathbf{w})$ is the partition function and the energy function E_s of our pairwise CRF model is formulated as follows:

$$E_s(\mathbf{x}, \mathbf{y}, \mathbf{w}) = \sum_{i=1}^{N} \phi_u(y_i, x_i, \mathbf{w}) + \alpha \sum_{(x_i, x_j) \in \mathcal{N}} \phi_p(y_i, y_j, x_i, x_j) \tag{2}$$

where ϕ_u and ϕ_p represent the unary term and pairwise term, respectively. Here, $\mathcal{N}$ denotes the set of spatially adjacent supervoxels. The parameter α controls the weight of the pairwise term. $\mathbf{w}$ is the parameters in the unary term ϕ_u.

The unary term $\phi_u(y_i, x_i, \mathbf{w})$ measures how well supervoxel x_i takes category y_i under current model $\mathbf{w}$. We define our unary term as follows:

$$\phi_u(y_i, x_i, \mathbf{w}) = -\log(P_u(y_i|x_i, \mathbf{w})) \tag{3}$$

where $P_u(y_i|x_i, \mathbf{w})$ is the probability of category label y_i taken by supervoxel x_i. To obtain P_u, given the features or descriptors of supervoxels, one-versus-all RF classifiers [31] are first learned for each category in a training set. Then, once the RF classifiers are learned, their probabilistic output, $P_r(y_i|x_i)$, of supervoxel x_i taking category y_i is calibrated via a multiclass logistic classifier [32] as follows:

$$P_u(y_i|x_i, \mathbf{w}) = \frac{1}{1 + \exp(w_a P_r(y_i|x_i) + w_b)} \tag{4}$$

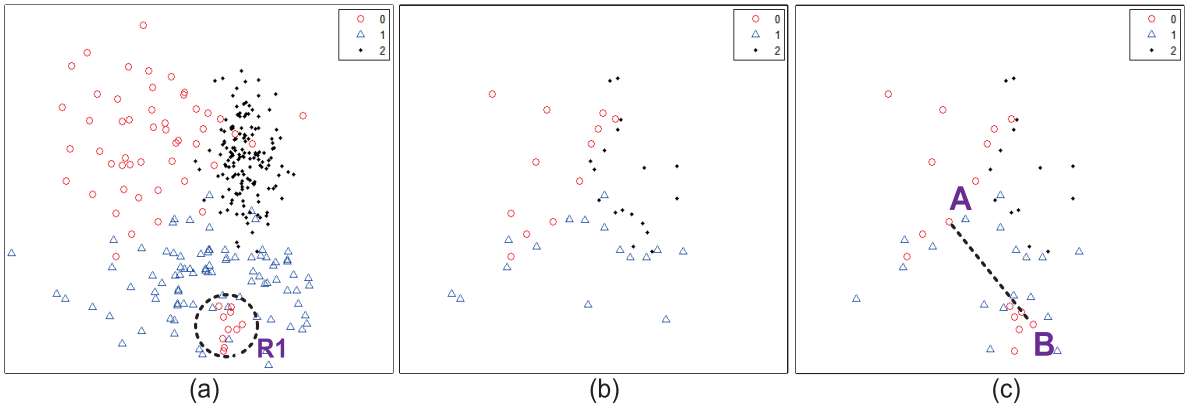


Fig. 3. Toy examples of active selection of samples. (a) Unlabeled sample pool. (b) Sample selection by MS. (c) Sample selection by MMbS considering the neighbor-consistency prior. In (c), the dotted line represents that samples A and B are spatially adjacent.

where w_a and w_b are the parameters of the sigmod function that is estimated using a maximum likelihood method for optimizing the training set. These parameters are obtained by a gradient descent search method.

The pairwise energy term ϕ_p takes the Potts model [24], which encourages neighboring supervoxels of similar feature with the same category. We define our pairwise term as follows:

$$\phi_p(y_i, y_j, x_i, x_j) = \begin{cases} D(x_i, x_j), & y_i \neq y_j \\ 0, & y_i = y_j \end{cases} \tag{5}$$

where $D(x_i, x_j)$ is a similarity metric which measures the similarity of two supervoxels. We scale the value of $D(x_i, x_j)$ to [0, 1] to meet the requirement of submodular. To this end, given the unary term and pairwise term, the labeling $\widehat{\mathbf{y}}$ can be predicted by efficiently minimizing the energy function (2) through the α-expansion algorithm [33]

$$\widehat{\mathbf{y}} = \underset{\mathbf{y} \in \mathcal{L}^N}{\operatorname{argmin}} E_s(\mathbf{x}, \mathbf{y}, \mathbf{w}). \tag{6}$$

2) Modified Margin-Based Sampling: The margin-based sampling (MS) [34], as a basic active learning algorithm, actively selects valuable samples to reduce the model uncertainty by focusing on the margins of current classifiers. The margin-based uncertainty, $\mathcal{MU}(x_i)$, of a supervoxel x_i is measured by (7), which computes the difference between best versus second best class prediction

$$\mathcal{MU}(x_i) = P_u\left(\hat{y}_i^2 | x_i, \mathbf{w}\right) - P_u\left(\hat{y}_i^1 | x_i, \mathbf{w}\right) \tag{7}$$

where $\hat{y}_i^1$ and $\hat{y}_i^2$ are the first and second most probable class labels under current statistical model, respectively. The higher value of $\mathcal{MU}(x_i)$ means that supervoxel x_i is more valuable and uncertain. Therefore, in MS, samples nearby the margins of classifiers are considered uncertain to a model.

As illustrated in Fig. 3, MS can effectively select samples nearby the margin of current classifiers, but ignore some sample distributions, e.g., the region $R1$, which are surrounded by other categories and away from the margin. However, those samples from these ignored distributions may be crucial for the learning procedure needed to train discriminative classifiers.

Commonly, samples in those ignored regions are misclassified by current model. Intuitively, samples from those ignored regions can be incorporated into training set by searching misclassified samples. In addition, from the perspective of classification, selecting the misclassified samples into training set assists in gradually improving the accuracy of classifiers. In order to find misclassified samples, the neighbor-consistency prior that pairwise supervoxels have a high probability of taking the same category label is considered into the sampling procedure [see Fig. 3(c)]. Here, pairwise supervoxels are defined as two spatially adjacent supervoxels.

Based on the neighbor-consistency prior, if one supervoxel x_j in pairwise supervoxels (x_i, x_j) with different categories has been known its true category label $\overline{y_j}$, we can define the misclassified possibility, $\mathcal{MP}(x_i)$, of supervoxel x_i as follows:

$$\mathcal{MP}(x_i) = 1 - P_u(\overline{y_j} | x_i, \mathbf{w}). \tag{8}$$

Equation (8) implies that higher misclassified probability will be given to supervoxel x_i, if the inferred category of supervoxel x_i has the lower probability of the category which is the same with the true category of its neighboring supervoxel x_j.

The MMbS is proposed by introducing the neighbor-consistency prior into MS (see Algorithm 2). The MMbS selects potentially misclassified samples to cover the ignored sample distributions while considering determination of accurate margins for classifiers. More concretely, in Algorithm 2, lines 1–4 apply the MS to sample the informative samples by focusing on the margins of classifiers. Based on the true categories of the samples selected by the MS, lines 4–9 exploit the neighbor-consistency prior to select the possibly misclassified samples. Threshold ρ allows us to select the samples with high misclassified probability.

3) Active Selection Procedure: As illustrated in Fig. 2, at each iteration of active learning, a pairwise CRF model is first learnt and updated over a set of manually annotated supervoxels. Second, the pairwise supervoxels with different inferred labels are collected. Third, only pairwise supervoxels containing minority category are taken as input to the MMbS. Here, the minority category is dynamically determined by the current set of manually annotated supervoxels. This strategy

Algorithm 2 Modified Margin Sampling to Actively Select Valuable Supervoxels

Input: a set of pairwise supervoxels $\mathcal{D} = \{(x_i, x_j)\}$ inferred with different categories

Output: the manually-annotated supervoxel set $\mathcal{D}_L^*$

```
for each supervoxel not inferred as minority category in D, compute MU by Eq. (7)
select the supervoxel x* with highest MU and obtain its true label y*
insert (x*, y*) into D_L^*
for each pairwise supervoxel, (x_i, x*), compute MP of supervoxel, x_i, by Eq. (8)
select the supervoxel, x_i', with highest MP
if MP(x_i') > ρ then
    obtain true label, y_i', of supervoxel, x_i',
    insert (x_i', y_i') into D_L^*
end if
return D_L^*
```

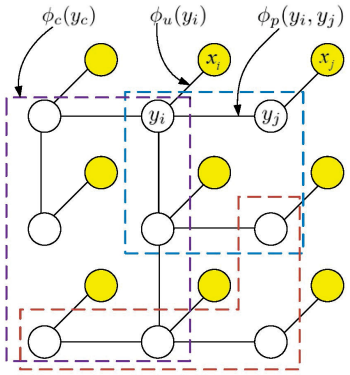


Fig. 4. Graphic example of a higher order MRF. ϕ_u represents unary potential, ϕ_c represents pairwise potential, and ϕ_p represents higher order potential.

assists in keeping diversity in the composition of the training set by avoiding the sampling procedure being trapped in one category. Finally, through MMbS algorithm, the most valuable supervoxels are selected and manually annotated.

All the above steps are performed in each iteration. Iterations terminate when a defined maximum iteration is reached. Once the iterations are terminated, a pairwise CRF model is finally trained based on manually annotated supervoxels for semantic labeling of mobile LiDAR point clouds.

D. Label Refinement by Higher Order MRF

As shown in Fig. 2, there is a portion of the inaccurate categorial labels in initial labeling results obtained by applying pairwise CRF. This is because only short-range energy term (pairwise energy potential) is insufficient to describe long-range interactions among the supervoxels from point cloud scenes. We propose to exploit higher order MRF to consider more contexts into label refinement. As shown in Fig. 4, pairwise potential only models the interaction between two variables. However, higher order potential can describe the interactions among variables belonging to a clique (region).

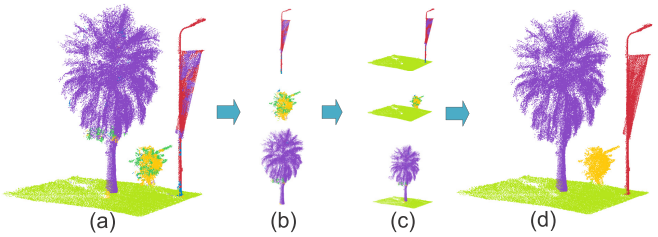


Fig. 5. Example of label refinement with regional label cost. (a) Initial labeling result obtained by applying a pairwise CRF model. (b) Regions generated by the clustering algorithm. (c) Final region used in label refinement. (d) Refined labeling with considering regional label cost.

Therefore, we design the energy function of the higher order MRF as follows:

$$E(\mathbf{y}) = E_u(\mathbf{y}) + \alpha E_p(\mathbf{y}) + \beta E_c(\mathbf{y}) \tag{9}$$

where α and β are the weights of pairwise term E_p and higher order term E_c, respectively. The unary term E_u and pairwise term E_c are defined as (2). In addition, the related parameters are set to be the same as the pairwise CRF trained in active learning procedure.

The higher order term E_c is designed by using the label cost term introduced in [35]. The label cost term tends to reduce redundant label categories by imposing the cost of these labels that exist in a category subset. In our proposed framework, the purpose of introducing a label cost term in our proposed framework is to use fewer category labels to describe a region in point cloud scenes by penalizing redundant categories (see Fig. 5). By eliminating the unnecessarily used categories in a region, many mislabeled points in initial labeling results may be rectified. We define the higher order term E_c as follows:

$$E_c(\mathbf{y}) = \sum_{r \in R} E^r_{\text{label}}(\mathbf{y}) \tag{10}$$

where R represents the region set in a point cloud scene. $E^r_{\text{label}}(\mathbf{y})$ represents the region r's label term which penalizes each unique label that appears in region r

$$E^r_{\text{label}}(\mathbf{y}) = \sum_{l \in \mathbf{y}} h_r(l) \cdot \delta_r(l) \tag{11}$$

where $h_r(l)$ is a nonnegative label cost of label l and is given by (13). $\delta_r(l)$ is a function that indicates whether label l is used in labeling region r

$$\delta_r(l) = \begin{cases} 1, & \exists x_i \in r : y_i = l \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

$$h_r(l) = \begin{cases} \exp\left(\dfrac{M_l - |S_r(l)|}{M_l}\right), & |S_r(l)| < M_l \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

$$S_r(l) = \{x_i | \forall x_i \in r \wedge \tilde{y}_i = l\} \tag{14}$$

where $S_r(l)$ represents the set of supervoxels, which belong to category l in region r. $\tilde{y}_i$ is the initial category label of supervoxel x_i. $|S|$ represents the size of set S. M_l is a constant number for a specific label l.

By using (13), the label term penalizes category l heavily when there are a few supervoxels labeled as category l. In addition, (13) also implies that in region r, if the number of supervoxels of a specific category l is larger than a constant number M_l, we will assume that the specific category l is in region r. Intuitively, M_l should be related to the size of objects in category l. Thus, in the experiments, we set M_l according to the number of supervoxels belonging to individual object of category l.

To impose constraints on category labels in a region, it is critical to define regions in a scene. In our framework, a clustering algorithm is carried out to generate regions through clustering adjacent supervoxels. In the clustering algorithm, the basic operational units are supervoxels with category labels, which are obtained by applying the trained pairwise CRF. Terminating the growth of a region should meet one of two conditions: 1) there is no supervoxel adjacent to the region and 2) all the supervoxels adjacent to the region should belong to termination regions. Here, a termination region is defined as a set of spatially connected supervoxels with same category labels, and the size of the set of connected supervoxels should be larger than a defined constant $\rho_{\max}$. In general, the easily classified categories, such as ground and grass, are used to define the termination regions. Once the growth of the region is terminated, a region [see Fig. 5(c)] used in the label refinement is defined by two parts: a region generated by the proposed clustering algorithm [see Fig. 5(b)] and its connected termination regions.

Once region extraction is completed, energy E is minimized by Algorithm 3 which iteratively implements the extending α-expansion algorithm introduced in [35]. Finally, the refined labeling results [see Fig. 5(d)] are obtained.

Algorithm 3 Label Refinement by Regional Label Costs

1: define the regions according to initial labeling
2: compute $h_r(l)$ and $S_r(l)$ for each defined region
3: for each region, re-estimate the labeling by extending α-expansion algorithm [35]

IV. Results and Discussion

To quantitatively evaluate the accuracy and correctness of the proposed method on semantic labeling of point clouds, three measurements, including precision, recall, and F1-measure [18], were selected. Precision describes the percentage of true positives in the ground truth; recall depicts the percentage of true positives in the semantic labeling results; and F1-measure is an overall measure. The three measurements are calculated on points and defined as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{15}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16}$$

$$\text{F1-measure} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{17}$$

where TP, FN, and FP represent the numbers of true positives, false negatives, and false positives, respectively.

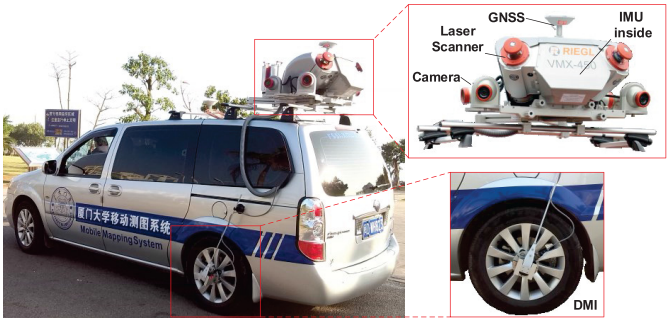


Fig. 6. Illustration of the REIGL VMX-450 mobile LiDAR system and its configurations.

A. Experimental Data Set

Devoted to illustrating the capabilities of our presented framework on semantic labeling of mobile LiDAR point clouds, we perform both qualitative and quantitative evaluations on three different data sets.

The point clouds in both data sets I and II are collected by an RIEGL VMX-450 mobile LIDAR system [36] on Xiamen Island, China. This LIDAR system, smoothly integrating two RIEGL VQ-450 laser scanners, a global navigation satellite system antenna, an inertial measurement unit, a distance measurement indicator, and four high-resolution digital cameras, was mounted on the roof of a minivan with an average speed of 40–50 km/h (Fig. 6). The point density of acquired points is about 7000 points/m^2. The accuracy and precision of the scanned point clouds are within 8 and 5 mm, respectively. After data acquisition, we used RiProcess, a postprocess software released by REIGL corporation, to obtain colorized mobile LiDAR point clouds by registering the images with point clouds. To evaluate the performance of semantic labeling methods, two data sets of road scenes are built by manually classifying all the points. Data set I consists of eight challenging categories: palm tree, cycas, brushwood, vehicle, light pole, grass, and road. Data set II contains seven challenging categories: tree, vehicle, wall, light pole, ground, and pedestrian. As shown in Table I, there is a category imbalance problem in both data sets, e.g., the points of light poles and vehicle are much fewer than the other categories (data set I); the points of light poles and pedestrian are much fewer than the other categories (data set II). In addition to challenges brought by category imbalances, other challenges, such as intraclass variations, interclass similarities, overlapping, and object incompleteness, commonly exist in our ground truth.

The point clouds in data set III are collected around CMU campus in Oakland, Pittsburgh, PA, USA, by using the Navlab11 equipped with side looking SICK LMS laser scanners. Due to lack of cameras in the Navlab11, there is no color information in the collected point clouds. Four categories (ground, building, vehicle, and trees) provided in [11] are used in our experiments. As shown in Table I, the amount of the points in data set III is much smaller than those in data sets I and II. This is because the point density in data set III is much lower than those in data sets I and II.

In our experimental setup, each data set is partitioned into two parts: training and testing samples (see Table I).

TABLE I
DESCRIPTION OF GROUND TRUTH (UNIT: K POINTS)

Dataset I	road	grass	palm tree	cycas	brushwood	light pole	vehicle	others
Training	10,120	3,154	1,021	423	300	130	51	30
Testing	80,441	25,275	7,110	4,188	2,875	527	428	175
Dataset II	**ground**	**tree**	**wall**	**vehicle**	**light pole**	**pedestrian**	**others**	
Training	9,261	5,212	474	596	34	32	9	
Testing	15,498	11,404	1,558	1,892	61	36	12	
Dataset III	**ground**	**tree**	**building**	**vehicle**				
Training	87	62	44	24				
Testing	688	185	62	39				

The training samples are used as forming the unlabeled sample pool for the active learning procedure. The testing samples are used to evaluate the performance of our proposed framework in labeling point clouds.

B. Manually Annotate Training Sets With Active Learning

In the pairwise CRF model used in active learning, for data sets I and II, we define the similarity metric $D(x_i, x_j)$ with (18). For data set III, we define the similarity metric $D(x_i, x_j)$ with (19) by using the χ^2 distance [37] of the FPFH descriptor of supervoxels x_i and x_j

$$D_{\text{color}}(x_i, x_j) = \exp\left(-\gamma \sum_{k=1}^{3} \frac{|C_i(k) - C_j(k)|}{255}\right) \quad (18)$$

$$D_{\text{fpfh}}(x_i, x_j) = \exp\left(-\gamma \sum_{k=1}^{16} \frac{[F_i(k) - F_j(k)]^2}{F_i(k) + F_j(k)}\right) \quad (19)$$

where $\mathbf{F}_i$ denotes a 16-D FPFH descriptor for a supervoxel x_i; $\mathbf{C}_i$ represents an RGB color vector of a supervoxel x_i; γ is a scale factor which makes the unary term and pairwise term comparable. In the experiments, we set γ at 15 to make unary term and pairwise term comparable.

In active learning, at each iteration, we use these manually annotated supervoxels as inputs to train a set of one-versus-all RF classifiers. The number of decision trees in the RF is set at 100. The depth of each tree is set at 15. Threshold ρ used in Algorithm 2 is set to 0.7. In the first iteration, the selected samples are initialized by randomly selecting 20 samples for each category to query their category labels. During the sampling procedure, as suggested in [38], we adopted the batch model, which selects multiple supervoxels to be annotated manually at each iteration, to reduce the overwhelming computational complexity brought by the serial model. More specifically, all the pairwise supervoxels, which are the inputs to Algorithm 2, are clustered into several groups by applying k-means clustering [39]. Five clusters are obtained according to the feature descriptors of the supervoxels which are not inferred as the current minority category. Then within each group, the MMbS is applied to select valuable supervoxels.

1) Qualitative and Quantitative Results: To assess the performance of the proposed MMbS in actively creating a promising and minimal training set, we perform both qualitative and quantitative evaluations on all the data sets. Initial

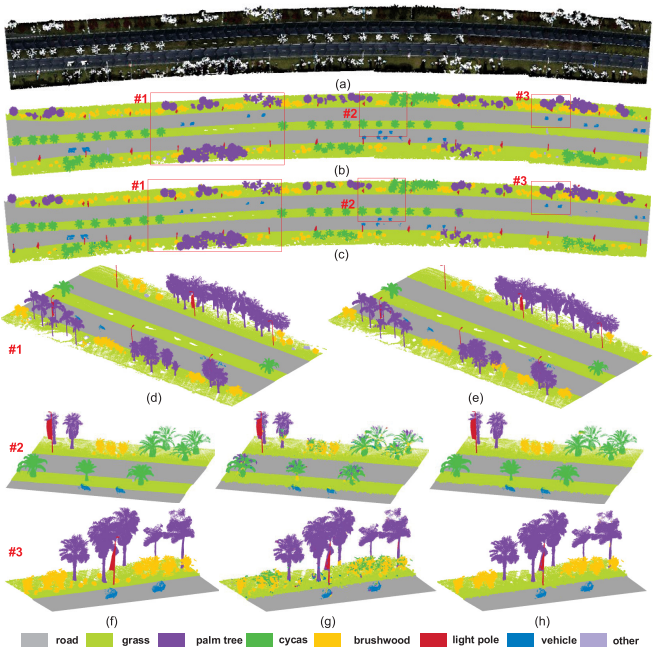


Fig. 7. Qualitative labeling results on a part of data set I. (a) Colorized point clouds. (b) Ground truth. (c) Semantic labeling results. (d) and (f) Close-up views of the ground truth in areas #1, #2, and #3. (g) Close-up views of the initial labeling results obtained by applying pairwise CRF model in areas #2 and #3. (e) and (h) Close-up views of the refined results obtained by incorporating regional label costs in areas #1, #2, and #3.

labeling results obtained by applying the pairwise CRF model are shown in Figs. 7(g), 8(c) and (d), and 9(e) and (f). Although there is a small portion of mislabeled points caused by local feature similarities, the majority of the points in the initial results are correctly classified, which prove the effectiveness of MMbS in our proposed framework. Moreover, as shown in Tables II–IV, the average initial labeling results (AL-Pairwise) achieved in precision, recall, and F1-measure on data sets I–III are (0.794, 0.69, 0.772), (0.818, 0.773, 0.781), and (0.879, 0.867, 0.873), respectively. The quantitative results demonstrate the feasibility of our proposed MMbS to create a small training set for training a labeling model which can perform well on classifying 3-D points.

2) Comparison With Traditional Active Learning Methods: To exhibit the superior performance of our proposed sampling method over other traditional active learning method, we compared the proposed MMbS with three competing

TABLE II
EXPERIMENTAL RESULTS OF DIFFERENT APPROACHES ON DATA SET I

		road	grass	palm tree	cycas	brushwood	light pole	vehicle	argv
Precision	Shape-based [41]	-	-	.907	.648	.136	.731	.56	.596
	M3N [10]	.971	.933	.891	.80	.538	.526	.65	.759
	3D-PMG + MRF [21]	.971	**.955**	**.96**	**.926**	.564	.742	.762	.84
	AL-Pairwise (ours)	.988	.933	.928	.565	.700	.731	.710	.794
	AL-Pairwise+LabelCost (ours)	**.987**	.966	.922	.895	**.680**	**.924**	**.903**	**.897**
Recall	Shape-based [41]	-	-	.773	.532	.455	**.958**	.598	.663
	M3N [10]	.988	.876	.867	.718	.624	.862	.852	.827
	3D-PMG + MRF [21]	.989	.883	.962	.793	.81	.881	**.956**	**.896**
	AL-Pairwise (ours)	.984	**.946**	.837	.780	.667	.769	.401	.769
	AL-Pairwise+LabelCost (ours)	**.991**	.933	**.964**	**.862**	**.826**	.934	.724	.890
F1-measure	Shape-based [41]	-	-	.835	.585	.209	.83	.578	.607
	M3N [10]	.98	.904	.879	.757	.578	.653	.737	.784
	3D-PMG + MRF [21]	.98	.918	**.961**	.855	.665	.805	**.847**	.862
	AL-Pairwise (ours)	.986	.940	.880	.655	.668	.750	.513	.772
	AL-Pairwise+LabelCost (ours)	**.989**	**.949**	.943	**.878**	**.746**	**.929**	.804	**.891**

TABLE III
EXPERIMENTAL RESULTS OF DIFFERENT APPROACHES ON DATA SET II

		ground	tree	wall	vehicle	light pole	pedestrian	argv
Precision	Shape-based [41]	-	-	-	-	-	-	-
	M3N [10]	**.996**	.981	.912	.901	.196	.006	.665
	3D-PMG + MRF [21]	.995	.996	.907	.933	.531	**.552**	.819
	AL-Pairwise (ours)	.994	.986	.944	.867	**.693**	.420	.818
	AL-Pairwise+LabelCost (ours)	.995	**.998**	**.963**	**.941**	.576	.464	**.823**
Recall	Shape-based [41]	-	-	-	-	-	-	-
	M3N [10]	**.997**	.979	.904	.872	.473	.143	.728
	3D-PMG + MRF [21]	.989	**.997**	**.980**	.961	.309	.223	.743
	AL-Pairwise (ours)	.992	.994	.895	.962	.666	.123	.773
	AL-Pairwise+LabelCost (ours)	.991	.996	.962	**.984**	**.917**	**.286**	**.856**
F1-measure	Shape-based [41]	-	-	-	-	-	-	-
	M3N [10]	**.996**	.980	.908	.886	.277	.011	.676
	3D-PMG + MRF [21]	.992	.996	.942	.947	.391	.318	.765
	AL-Pairwise (ours)	.993	.990	.919	.913	.679	.190	.781
	AL-Pairwise+LabelCost (ours)	.993	**.997**	**.963**	**.962**	**.707**	**.353**	**.829**

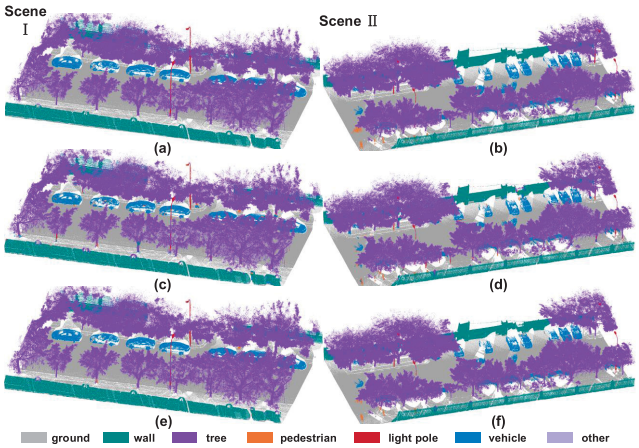


Fig. 8. Qualitative labeling results on two scenes in data set II. (a) and (b) Ground truth. (c) and (d) Initial labeling results. (d) and (e) Refined labeling results.

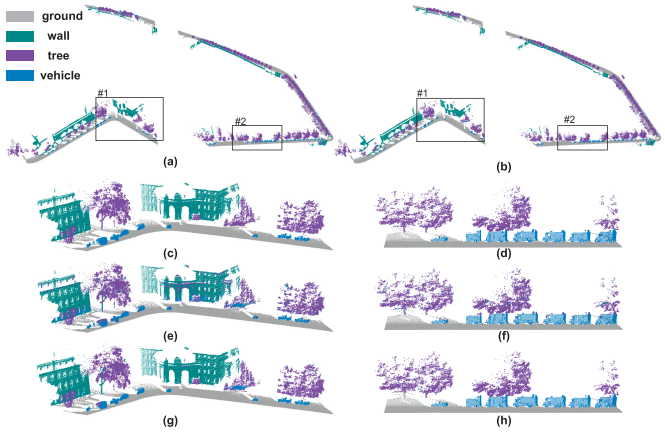


Fig. 9. Qualitative labeling results on data set III. (a) Ground truth. (b) Semantic labeling results. (c) and (d) Close-up views of the ground truth in areas #1 and #2. (e) and (f) Close-up views of the initial labeling results obtained by applying a pairwise CRF model in areas #1 and #2. (g) and (h) Close-up views of the refined results obtained by incorporating regional label cost areas #1 and #2.

sampling strategies: 1) a baseline random sampling (RS); 2) MS computed by (7); and 3) entropy-based sampling (ES) [40] computed by

$$\mathrm{Ent}(x_i) = -\sum_{y_i \in \mathcal{L}} P_u(y_i) \log(P_u(y_i)). \tag{20}$$

In order to compare the performance of sample selections conveniently, the label refinement step is not included in our comparative experiments. To eliminate the influence of

TABLE IV
EXPERIMENTAL RESULTS OF DIFFERENT APPROACHES ON DATA SET III

		ground	tree	vehicle	building	argv
Precision	[22]	**.999**	.875	.536	.789	.780
	3D-PMG + MRF [21]	.950	.963	**.997**	.768	.919
	AL-Pairwise (ours)	.996	.928	.783	.808	.879
	AL-Pairwise+LabelCost (ours)	.978	**.997**	.835	**.977**	**.947**
Recall	[22]	.971	.863	.772	.895	.875
	3D-PMG + MRF [21]	.959	.924	**.995**	.882	.940
	AL-Pairwise (ours)	**.997**	.936	.712	.822	.867
	AL-Pairwise+LabelCost (ours)	.987	**.992**	.921	**.945**	**.962**
F1-measure	[22]	.984	.869	.633	.839	.831
	3D-PMG + MRF [21]	.955	.943	**.996**	.821	.928
	AL-Pairwise (ours)	**.996**	.932	.746	.815	.873
	AL-Pairwise+LabelCost (ours)	.983	**.994**	.876	**.961**	**.954**

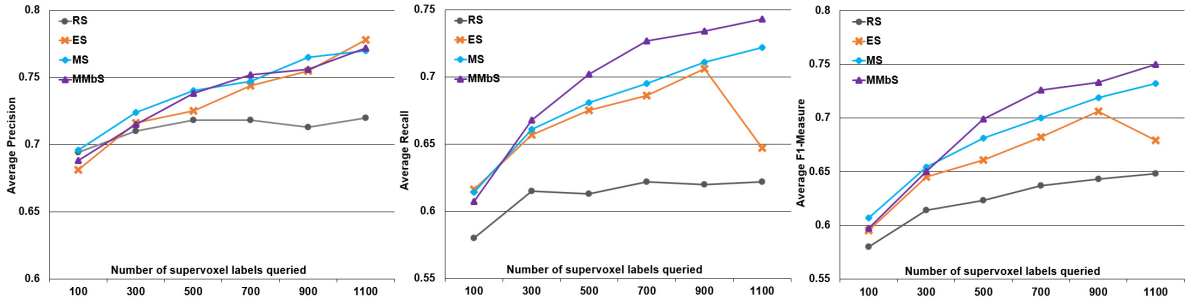


Fig. 10. Average labeling results achieved by the MMbS, ES, MS, and RS on data set I: average precision, average recall, and average F1-measure.

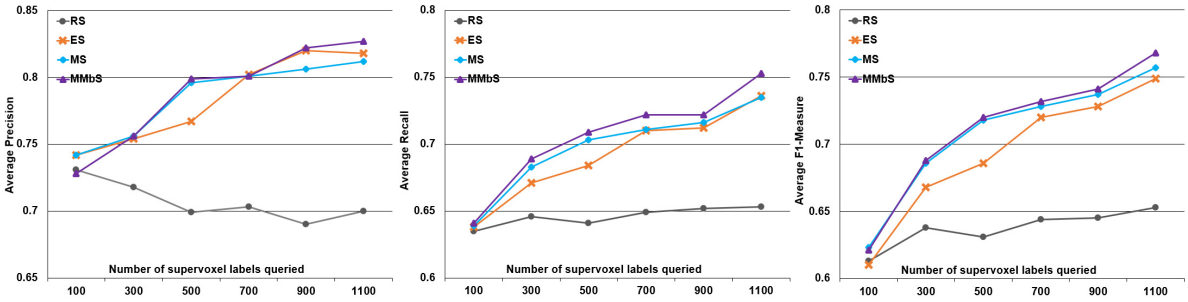


Fig. 11. Average labeling results achieved by the MMbS, ES, MS, and RS on data set II: average precision, average recall, and average F1-measure.

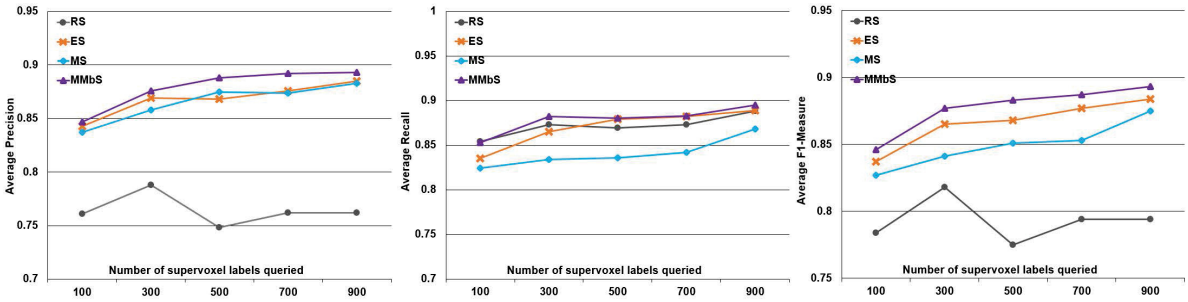


Fig. 12. Average labeling results achieved by the MMbS, ES, MS, and RS on data set III: average precision, average recall, and average F1-measure.

random initialization of annotated supervoxels, we repeated each sampling strategy 50 times. The mean values of each sampling method for average precision, recall, and F1-measure are recorded at different amounts of manually annotated samples.

The mean values for average precision, recall, and F1-measure on data sets I–III are shown in Figs. 10–12, respectively. As the number of supervoxel labels queried increases, the MMbS curves of precision, recall, and F1-measure demonstrate the stable performance of our

proposed sampling method. In addition, although the precisions of MMbS, ES, and MS are close (see Figs. 10 and 11), the curves of recall and F1-measure clearly demonstrate the superiority of our MMbS over other sampling methods, which reflects the effectiveness of exploiting neighbor-consistency prior to select potentially misclassified supervoxels into training sets.

C. Semantic Labeling With Higher Order MRF

At the label inferring stage, all the parameters used in (9) and (13) are experientially defined by visual inspection of the effect of the labeling results, and their values in our experiment settings are listed in Table V. In addition, during the region extraction procedure, the categories used in generating termination regions on data sets I–III are (road and grass), (ground and trees), and (ground), respectively. In addition, the constant $\rho_{\max}$ is set at 300.

1) Qualitative and Quantitative Results: To assess the performance of the proposed higher order MRF on refining the initial labeling results obtained by pairwise CRF, we exhibit both qualitative and quantitative evaluations on all built data sets. As presented in Figs. 7(h), 8(e) and (f), and 9(g) and (h), the refined labeling results demonstrate the promising capabilities of our proposed framework on labeling point clouds. Compared to the initial labeling results, a remarkable improvement was achieved. This is because the proposed higher order MRF can obtain smooth labelings by reducing redundant categories in a defined region. As the quantitative results reported in Tables II–IV, the average precision, recall, and F1-measure achieved by our proposed framework (AL-Pairwise+LabelCost) further demonstrate the proposed higher order MRF which reduces the redundant categories can help us to correct some misclassified points.

In addition, our proposed higher order MRF can effectively avoid oversmoothing overlapped objects and preserve overlapped objects. Therefore, the proposed higher order MRF performs well in the complex scenarios of overlapped objects. As shown in Figs. 7(h) and 8(e), although the tree and light poles are overlapped, our proposed higher order MRF avoid light poles being misclassified as tree, which shows the capabilities to deal with the scenario where objects overlapped. However, we find that a very small object may be mislabeled as its connected category. As shown in Fig. 7(h), small brushwood is oversmoothed and mislabeled as grass by our proposed higher order MRF.

As shown in Fig. 8(f), by considering the long-range contexts, our proposed higher order MRF correctly recognizes the moving and stationary vehicles, which shows its capability to handle the incompleteness and intraclass variations. However, as shown in Fig. 8(e) and (f), there are some tree trunks mislabeled as pedestrians; this is because in the initial labeling, the accuracy is low, and many points of a tree trunk are mislabeled as a pedestrian. Under these circumstances, the higher order MRF cannot rectify the mislabeled points.

D. Comparative Studies

To show the superior performance of our proposed framework in the semantic labeling of mobile LiDAR point clouds, the following three approaches were evaluated on data sets I and II for comparison: shape based [41], M3N [10], and 3-D-PMG based (3D-PMG+MRF) [21]. The settings of those approaches are the same as [21]. The shape-based approach tries to segment objects out of the point cloud scenes and then uses global features to recognize objects [41]. As shown by the quantitative results in Table II, the poor performance of the shape-based approach demonstrates that overlapping and incomplete objects in these complex scenarios are huge obstacles stymieing the success of these methods which depend on segmenting objects out of the whole scene. As shown in Table II, the performance of our AL-Pairwise achieves a lower average F1-measure than that of M3N approach whose average F1-measure is 0.784, because the M3N approach adopts a high-order potential energy term (a robust Potts model [25]) to model relatively long-range interactions among points. In addition, the 3D-PMG+MRF outperforms AL-Pairwise because of the consideration of object intrinsic and contextual properties when conducting label transfer. However, by exploiting the long-range contextual information, the AL-Pairwise+LabelCost approach, imposing regional label costs constraints on the initial labeling of AL-Pairwise, obtains better results than those of the other methods. Because the AL-Pairwise+LabelCost approach models not only short-range but also long-range contexts, it achieves a smoother labeling than that of 3D-PMG+MRF. As illustrated by the qualitative comparisons in Fig. 13, AL-Pairwise+LabelCosts preserve the vehicle, light poles, and palm trees better than those of 3D-PMG+MRF.

To further demonstrate the superiority of our proposed method on data set III, we conduct comparisons with the two following works: [21] and [22]. From Table IV, it is noted that our proposed method achieves the best results on data set III.

As illustrated in Table III, the AL-Pairwise outperforms the M3N and 3D-PMG+MRF on the data set II where scenarios are cluttered and more complex than data set I. This is because

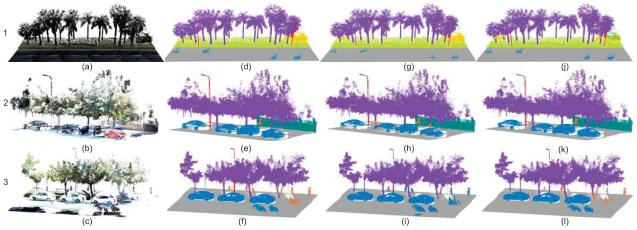


Fig. 13. Comparative labeling results on different scenes. (a)–(c) Colorized point clouds. (d)–(f) Ground truth. (g) and (h) Labeling results by applying 3D-PMG+MRF. (i) Labeling results by applying M3N. (j)–(l) Labeling results by applying our AL-Pairwise+LabelCosts.

TABLE V
PARAMETERS IN THE PROPOSED FRAMEWORK

α	10	β	60	M_{road}	300
M_{grass}	300	M_{palm}	250	$M_{brushwood}$	200
M_{cycas}	60	M_{pole}	10	$M_{vehicle}$	20
M_{wall}	300	$M_{pedestrian}$	5	M_{tree}	100
M_{ground}	300				

TABLE VI
TRAINING TIME ON DIFFERENT APPROACHES (UNITS: HOURS)

	Dataset I	Dataset II	Dataset III
Shape-based [41]	0.8	-	-
M3N [10]	2.4	1.7	-
3D-PMG + MRF [21]	5.5	3.2	2.8
AL-Pairwise+LabelCost (ours)	2.5	2.1	1.9

TABLE VII
LABELING TIME ON DIFFERENT APPROACHES (UNITS: HOURS)

	Dataset I	Dataset II	Dataset III
Shape-based [41]	4.2	-	-
M3N [10]	5.8	2.2	-
3D-PMG + MRF [21]	8.6	4.5	2.6
AL-Pairwise+LabelCost (ours)	4	2.3	2.5

complex and cluttered scenes cannot be well modeled for 3D-PMG+MRF, and M3N is not designed for the imbalanced data set. As reflected in Fig. 13(h), 3D-PMG+MRF mislabeled wall and vehicles to some extent because of the inaccurate color information caused by complex scenes. Thus, the AL-Pairwise+LabelCost modeling the higher order contexts obtains a more satisfied result [see Fig. 13(k)]. As reflected in Fig. 13(i), the M3N classified the light pole with many false positives and can hardly recognize pedestrians, while our proposed methods can correctly annotate pedestrians to some extent [see Fig. 13(l)]. The reason is that our sampling method exploits the neighbor-consistent prior to reduce the classification errors for the minority categories.

The proposed framework and comparative studies were coded with C++ and executed on a personal computer with a single Intel core of 3.30 GHz and a RAM of 16 GB. The processing time of the experiments was reported in Tables VI and VII. For our proposed framework, the training time containing the active learning procedure was approximate 2.5, 2.1, and 1.9 h, respectively, on three data sets. In addition, the labeling time on three data sets was 4, 2.3 and 2.5 h, respectively. The labeling times of our labeling framework are lower than those of shape-based, M3N, and 3D-PMG+MRF methods. Therefore, our proposed method has time–cost advantages.

From Figs. 7–9 and Tables II–IV, we can conclude that the presented framework can well distinguish the object classes from the point cloud of complex urban environments. The long-range contextual information encoded by higher order MRF can help us to correct some certain mislabeled classes and improves the labeling accuracy. Moreover, the proposed active learning method also assists in improving the classification accuracy by selecting the valuable samples to form a minimal training set.

E. Sensitivity of Proposed Framework

Here, we analyze the impact of the weight of regional label costs β on the performance of labeling mobile LiDAR point clouds. The analysis was performed on data set I. As reflected in Fig. 14, the F1-measure changes with the increase in parameter β, and these F1-measures obtained by all

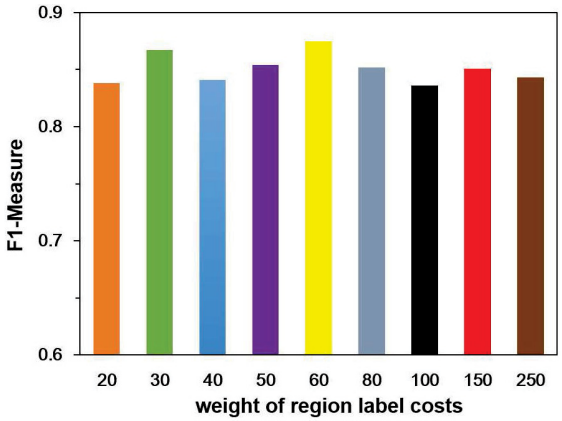


Fig. 14. Impact of the weight of regional label costs on semantic labeling results.

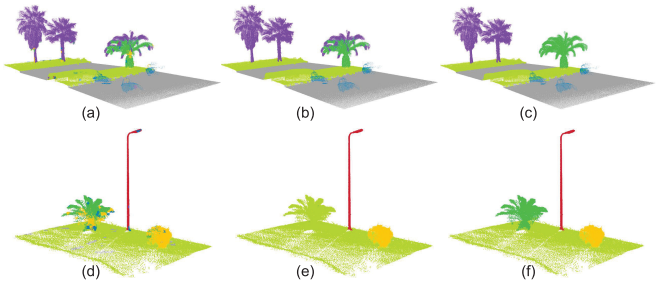


Fig. 15. Qualitative labeling results on two example scenarios with setting different weights β of regional label costs. (a) and (d) Initial labeling results. (b) Refined labeling results at $\beta = 20$. (c) and (f) Refined labeling results at $\beta = 60$. (e) Refined labeling results at $\beta = 120$.

these parameters show the improvement of the initial labeling results. Because a larger value of β means more costs imposed on the number of used categories, the F1-measure peak value is reached at a median value $\beta = 60$. The large costs may cause oversmooth labeling results, whereas a smaller β means fewer costs imposed on the number of used categories. Small costs may be inadequate to rectify a relatively large quantity of inaccurate labels. To further explain the influence of β, two example labelings given in Fig. 15 are used to illustrate the large and small cost scenarios, respectively. As shown in Fig. 15(b), the configuration of $\beta = 20$ in our proposed framework is too small to rectify the inaccurate labels of *cycas*. As reflected in Fig. 15(e), the configuration of $\beta = 120$ in our proposed framework is too big to preserve the accurately labeled objects cycas. The proper value of $\beta = 60$ achieves a promising refined labeling results [see Fig. 15(c) and (f)]. Therefore, to make a balance between the aforementioned two scenarios, we set the weight of regional label costs at $\beta = 60$.

To analysis the impact of number of queried supervoxels on the label refinement, both the initial and refined labeling results were recorded at the following configurations: 100, 300, 500, 700, 900, 1100, and 1300. As reflected in Fig. 16, all the curves going up with an increase of queried supervoxels demonstrate the stability of our framework. The curves of AL-Pairwise+LabelCost lie above the curves of AL-Pairwise in both two data sets. This is because our higher order potentials can rectify the mislabeled points to some extent. In addition, as the number of supervoxel labels queried

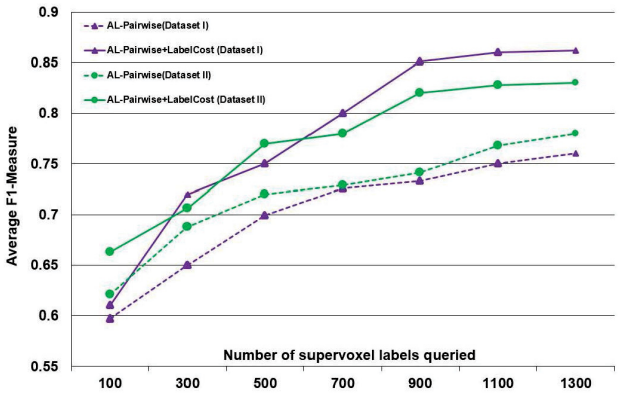


Fig. 16. Impact of the number of queried supervoxels on the refined labeling results.

increases from 900 to 1300, the average F1-measure values of AL-Pairwise+LabelCost increase slightly, whereas the average F1-measure values of AL-Pairwise increase. This shows that if the accuracy of labeling results has reached at a high value, the refined results will stay at a high value of F1-measure even though the initial results do not have a significant improvement. In our experiments, we set the number of queried supervoxels at 1100.

V. Conclusion

In this paper, we have presented a new framework which integrates active learning and higher order MRF for effectively conducting semantic labeling of mobile LiDAR point clouds. In order to manually annotate the 3-D point cloud data as small as possible, we introduce neighbor-consistency prior into active learning to select the potentially misclassified samples into training sets effectively. To consider more contexts into refining the labeling results, a higher order MRF encoding label cost terms is used to describe long-range interactions among supervoxels in a region. Quantitative evaluations on three different point cloud data sets have demonstrated that the proposed algorithm achieves average F1-measure of 0.891, 0.829, and 0.954, respectively. By considering long-range contextual information with higher order MRF, improvements of average F1-measure over the initial labeling results are up to 11.9%, 4.8%, and 8.1%, respectively, on three data sets. Comparative studies have also demonstrated that the proposed framework outperforms other traditional active learning methods in creating an optimal training set and other fully supervised semantic labeling methods in labeling point clouds. In conclusion, the proposed method is feasible and achieves satisfied performance in semantic labeling of mobile LiDAR point clouds with a small portion of manually annotated 3-D points.

References

[1] M. Vallati, D. Magazzeni, B. De Schutter, L. Chrpa, and T. L. McCluskey, "Efficient macroscopic urban traffic models for reducing congestion: A PDDL+ planning approach," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 3188–3194.

[2] M. A. S. Kamal, J.-I. Imura, T. Hayakawa, A. Ohata, and K. Aihara, "Smart driving of a vehicle using model predictive control for improving traffic flow," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 878–888, Apr. 2014.

[3] K. Mandal, A. Sen, A. Chakraborty, S. Roy, S. Batabyal, and S. Bandyopadhyay, "Road traffic congestion monitoring and measurement using active rfid and gsm technology," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2011, pp. 1375–1379.

[4] Y. Yu, J. Li, C. Wen, H. Guan, H. Luo, and C. Wang, "Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data," *ISPRS J. Photogram. Remote Sens.*, vol. 113, pp. 106–123, Mar. 2016.

[5] E. Levinkov and M. Fritz, "Sequential Bayesian model update under structured scene prior for semantic road scenes labeling," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1321–1328.

[6] J. Xiao and L. Quan, "Multiple view semantic segmentation for street view images," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 686–693.

[7] Z. Wang *et al.*, "A multiscale and hierarchical feature extraction method for terrestrial laser scanning point cloud classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2409–2425, May 2015.

[8] D. Anguelov *et al.*, "Discriminative learning of Markov random fields for segmentation of 3D scan data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 169–176.

[9] A. Nguyen and B. Le, "Contextual labeling 3D point clouds with conditional random fields," in *Intelligent Information and Database Systems*. Cham, Switzerland: Springer, 2014, pp. 581–590.

[10] D. Munoz, J. A. Bagnell, N. Vandapel, and M. Hebert, "Contextual classification with functional max-margin Markov networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 975–982.

[11] D. Munoz, N. Vandapel, and M. Hebert, "Onboard contextual classification of 3-D point clouds with learned high-order Markov random fields," in *Proc. IEEE Conf. Robot. Autom.*, May 2009, pp. 2009–2016.

[12] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, Jul. 2009.

[13] Y. Yang, Z. Ma, F. Nie, X. Chang, and A. G. Hauptmann, "Multi-class active learning by uncertainty sampling with diversity maximization," *Int. J. Comput. Vis.*, vol. 113, no. 2, pp. 113–127, 2015.

[14] Z. Wang and J. Ye, "Querying discriminative and representative samples for batch mode active learning," *ACM Trans. Knowl. Discovery Data*, vol. 9, no. 3, 2015, Art. no. 17.

[15] S. Z. Li, *Markov Random Field Modeling in Computer Vision*. Japan: Springer, 2012.

[16] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. Int. Conf. Mach. Learn.*, Jun./Jul. 2001, pp. 282–289.

[17] Z. Li *et al.*, "A three-step approach for TLS point cloud classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5412–5424, Sep. 2016.

[18] M. Najafi, S. T. Namin, M. Salzmann, and L. Petersson, "Non-associative higher-order Markov networks for point cloud classification," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 500–515.

[19] R. Shapovalov and A. Velizhev, "Cutting-plane training of non-associative Markov network for 3D point cloud segmentation," in *Proc. Int. Conf. 3D Imag., Modeling, Process., Vis. Transmiss. (3DIMPVT)*, May 2011, pp. 1–8.

[20] E. H. Lim and D. Suter, "3D terrestrial LIDAR classifications with super-voxels and multi-scale conditional random fields," *Comput.-Aided Des.*, vol. 41, no. 10, pp. 701–710, 2009.

[21] H. Luo *et al.*, "Patch-based semantic labeling of road scene using colorized mobile LiDAR point clouds," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 5, pp. 1286–1297, May 2015.

[22] Z. Li, L. Zhang, R. Zhong, T. Fang, L. Zhang, and Z. Zhang, "Classification of urban point clouds: A robust supervised approach with automatically generating training data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 3, pp. 1207–1220, Mar. 2017.

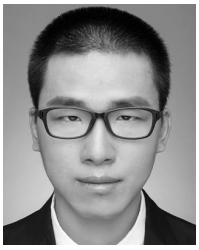
[23] H. Zhang, J. Wang, T. Fang, and L. Quan, "Joint segmentation of images and scanned point cloud in large-scale street scenes with low-annotation cost," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4763–4772, Nov. 2014.

[24] P. Kohli, M. P. Kumar, and P. H. S. Torr, "P3 & beyond: Solving energies with higher order cliques," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[25] P. Kohli, L. Ladický, and P. H. S. Torr, "Robust higher order potentials for enforcing label consistency," *Int. J. Comput. Vis.*, vol. 82, no. 3, pp. 302–324, May 2009.

[26] A. Vezhnevets, J. M. Buhmann, and V. Ferrari, "Active learning for semantic segmentation with expected change," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3162–3169.

[27] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.
[28] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, Jul. 2005.
[29] J. Papon, A. Abramov, M. Schoeler, and F. Wörgötter, "Voxel cloud connectivity segmentation—Supervoxels for point clouds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2027–2034.
[30] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. IEEE Conf. Robot. Autom.*, May 2009, pp. 3212–3217.
[31] A. Liaw and M. Wiener, "Classification and regression by random forest," *R Newslett.*, vol. 2, no. 3, pp. 18–22, 2002.
[32] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Adv. Large Margin Classifiers*, vol. 10, no. 3, pp. 61–74, 1999.
[33] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
[34] T. Luo *et al.*, "Active learning to recognize multiple types of plankton," *J. Mach. Learn. Res.*, vol. 6, pp. 589–613, Apr. 2005.
[35] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov, "Fast approximate energy minimization with label costs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2173–2180.
[36] *RIEGL VMX-450 Datasheet*. Accessed: May 2, 2015. [Online]. Available: http://www.riegl.com/nc/products/mobile-scanning/produktdetail/product/scannersystem/10/
[37] E. Wahl, U. Hillenbrand, and G. Hirzinger, "Surflet-pair-relation histograms: A statistical 3D-shape representation for rapid classification," in *Proc. 4th Int. Conf. 3-D Digit. Imag. Modeling*, Oct. 2003, pp. 474–481.
[38] X.-Y. Zhang, S. Wang, and X. Yun, "Bidirectional active learning: A two-way exploration into unlabeled and labeled data set," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3034–3044, Dec. 2015.
[39] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient K-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.
[40] A. J. Joshi, F. Porikli, and N. Papanikolopoulos, "Multi-class active learning for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2372–2379.
[41] A. Golovinskiy, V. G. Kim, and T. Funkhouser, "Shape-based recognition of 3D point clouds in urban environments," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2154–2161.



Huan Luo received the B.Sc. degree in software engineering from Nanchang University, Nanchang, China, in 2009, and the Ph.D. degree in computer science from Xiamen University, Xiamen, China, in 2017.

He is currently a Faculty Member with the College of Mathematics and Computer Science, Fuzhou University, Fuzhou, China. His research interests include point cloud processing, computer vision, and machine learning.

Cheng Wang (M'11–SM'16) received the Ph.D. degree in communication and signal processing from the National University of Defense Technology, Changsha, China, in 2002.

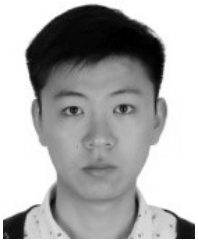
He is currently a Professor with the School of Information Science and Engineering and the Executive Director of the Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, Xiamen, China. He has co-authored over 150 papers in referred journals and top conferences including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Pattern Recognition*, IEEE-TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, the Association for the Advancement of Artificial Intelligence conference, and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing image processing, point cloud analyses, multisensor fusion, and mobile mapping.

Dr. Wang is the Chair of the ISPRS Working Group I/6 on multisensor integration and fusion. He is also a Fellow of The Institution of Engineering and Technology.

Chenglu Wen (M'14–SM'17) received the Ph.D. degree in mechanical engineering from China Agricultural University, Beijing, China, in 2009.

She is currently an Associate Professor with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Information Science and Engineering, Xiamen University, Xiamen, China. She has co-authored over 30 research papers in refereed journals and proceedings. Her research interests include machine vision, machine learning, and point cloud processing.

Dr. Wen is the Secretary of the ISPRS Working Group I/3 on multiplatform multisensor system calibration during 2012–2016.



Ziyi Chen received the Ph.D. degree in signal and information processing from Xiamen University, Xiamen, China, in 2016.

He is currently a Lecturer with the Department of Computer Science and Technology, Huaqiao University, Quanzhou, China. His research interests include computer vision, machine learning, and remote sensing image processing.

Dawei Zai received the B.S. degree in aircraft design and engineering from Xian Jiaotong University, Xi'an, China. He is currently pursuing the Ph.D. degree with the Department of Communication Engineering, Xiamen University.

His research interests include 3-D vision, machine learning, and point cloud data processing.

Yongtao Yu received the B.Sc. and Ph.D. degrees in computer science and technology from Xiamen University, Xiamen, China, in 2010 and 2015, respectively.

He is currently an Assistant Professor with the Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Nanjing, China. He has co-authored over 20 research papers in refereed journals and proceedings. His research interests include pattern recognition, computer vision, machine learning, intelligent interpretation of point clouds, and remotely sensed imageries.

Jonathan Li (M'00–SM'11) received the Ph.D. degree in geomatics engineering from the University of Cape Town, Cape Town, South Africa.

He is currently a Professor and the Head of the Mobile Sensing and Geodata Science Lab, Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada. He has co-authored over 350 publications, over 150 of which were published in refereed journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *ISPRS Journal of Photogrammetry and Remote Sensing, International Journal of Remote Sensing, Photogrammetric Engineering & Remote Sensing*, and *Remote Sensing of Environment*. His research interests include information extraction from Light Detection and Ranging (LiDAR) point clouds and from earth observation images.

Dr. Li is the Chair of the ISPRS Working Group I/2 on LiDAR for Airborne and Spaceborne Sensing during 2016–2020, and the ICA Commission on Sensor-driven Mapping during 2015–2019, and an Associate Editor of the IEEE-TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.