A Multilevel Point-Cluster-Based Discriminative Feature for ALS Point Cloud Classification

Zhenxin Zhang, Liqiang Zhang, Xiaohua Tong, P. Takis Mathiopoulos, *Senior Member, IEEE*, Bo Guo, Xianfeng Huang, Zhen Wang, and Yuebin Wang

***Abstract*—Point cloud classification plays a critical role in point cloud processing and analysis. Accurately classifying objects on the ground in urban environments from airborne laser scanning (ALS) point clouds is a challenge because of their large variety, complex geometries, and visual appearances. In this paper, a novel framework is presented for effectively extracting the shape features of objects from an ALS point cloud, and then, it is used to classify large and small objects in a point cloud. In the framework, the point cloud is split into hierarchical clusters of different sizes based on a natural exponential function threshold. Then, to take advantage of hierarchical point cluster correlations, latent Dirichlet allocation and sparse coding are jointly performed to extract and encode the shape features of the multilevel point clusters. The features at different levels are used to capture information on the shapes of objects of different sizes. This way, robust and discriminative shape features of the objects can be identified, and thus, the precision of the classification is significantly improved, particularly for small objects.**

***Index Terms*—Latent Dirichlet allocation (LDA), point cloud classification, point-cluster-based features, sparse coding.**

I. Introduction

THE airborne laser scanning (ALS) technique provides accurate 3-D coordinates of large landscapes and multiecho pulses and intensities with high horizontal and perpendicular accuracy [1]. For applications such as 3-D urban modeling and city planning, efficient representation and recognition of urban scenes from ALS point clouds are highly desirable. However, accurately classifying objects in complex landscapes is a difficult task [2] due to their large variety, complex geometries, and visual appearances. Extracting the discriminative features of objects from noisy and cluttered scenes is a key step in producing excellent classification results. ALS devices are usually far from the objects, and therefore, the number of points sampled is different for different objects, i.e., the larger the size of the object facing the device is, the more points are sampled, and vice versa. Traditionally, range scan processing algorithms have focused either on small objects in isolation or on large objects in scenes. Simultaneously extracting the useful shape features of large and small objects from ALS point clouds remains a challenge.

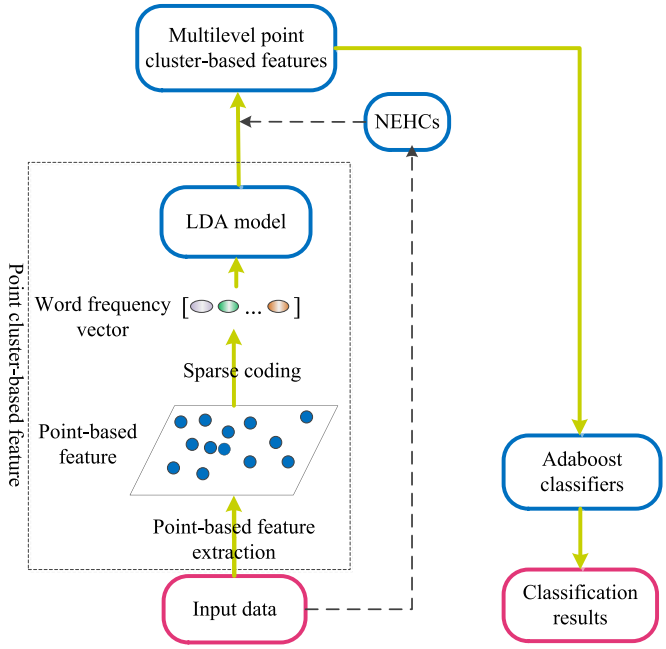


Fig. 1. Process of classifying ALS point clouds.

Our main goal is to present a novel framework for effectively extracting the shape features of objects from ALS point clouds that is based on hierarchical point clusters. The framework is then used to classify small and large objects in point clouds. As illustrated in Fig. 1, the proposed method first aggregates the input ALS point cloud into multilevel point-cluster sets based on natural exponential function thresholds (NEHCs). Then, joint latent Dirichlet allocation and sparse coding (SCLDA) is performed to construct the shape features of the objects using the point-based features. Finally, the AdaBoost classifier is used to classify on-ground objects from complex ALS point

Manuscript received October 30, 2015; revised December 18, 2015; accepted December 31, 2015. Date of publication January 20, 2016; date of current version April 27, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 41371324, Grant 41325005, and Grant 41501499 and in part by the Shenzhen Scientific Research and Development Funding Program under Grant JCYJ20150625102531697.

Z. Zhang, L. Zhang, and Y. Wang are with the State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China (e-mail: zhenxin066@163.com; zhanglq@bnu.edu.cn; xxgcdxwyb@163.com).

X. Tong is with the School of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: xhtong@tongji.edu.cn).

P. T. Mathiopoulos is with the Department of Informatics and Telecommunications, National and Kapodestrian University of Athens, Athens 157 84, Greece (e-mail: mathio@di.uoa.gr).

B. Guo is with the Key Laboratory for Geo-Environment Monitoring of Coastal Zone of National Administration of Surveying, Mapping and Geoinformation and the Shenzhen Key Laboratory of Spatial Smart Sensing, Shenzhen University, Shenzhen 518060, China (e-mail: guobo.szu@qq.com).

X. Huang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: huangxf@whu.edu.cn).

Z. Wang was with the State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China. He is now with the Department of Geoinformatics, China University of Geosciences, Beijing, China (e-mail: comige@gmail.com).

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.2016.2514508

clouds. Our method is then tested on ALS point clouds acquired from different urban scenes. The classification precision of our method is better than that of other methods, particularly in terms of classifying small objects, such as cars, and our method has a significant advantage over other methods.

The main contributions of this paper are as follows.

1) A natural exponential function threshold-based approach is proposed to construct hierarchical point clusters, i.e., NEHCs. The input point cloud is split into hierarchical point clusters with different sizes based on the natural exponential function threshold. The point cloud is split into hierarchical point clusters based on the sizes of the on-ground objects and the characteristics of the data. Therefore, the features extracted from the NEHCs are more robust than the point-based features, and they can be adapted to point clouds with various point densities, which is particularly important for small on-ground objects such as cars.
2) SCLDA, which integrates the advantages of latent Dirichlet allocation (LDA) and sparse coding, is used to extract the shape features of NEHCs. LDA can extract discriminative and robust features that have the same size from point clusters containing different numbers of points, whereas sparse coding can identify correlations between different features, eliminate redundant features, and retain discriminative features. By virtue of SCLDA inheritance, discriminative information for multiple levels can be leveraged to further improve the separability of low-level points.

II. Related Work

A. Feature Representation

As noted in [3], the generation of good features is very important for obtaining high-precision classification results. The features are generally divided into point-based [4], [5], voxel-based [6]–[9], and object-based [10] levels.

In the study of point-based feature classification, Niemeyer *et al.* [4] first derived features from full-waveform LiDAR point cloud data and then used a conditional random field (CRF) to integrate the point-based features with contextual information for pointwise classification. In [5], the point cluster features were developed using a multiscale and hierarchical point cloud partition framework.

Unlike point-level classification features, scene analysis may be also based on voxels or 3-D segments. Lim and Suter [6] combined generic scale selection and voxelization to iteratively oversegment a point cloud into supervoxels and then computed the local and regional features of the supervoxels. They later [7] showed that, although supervoxels can reduce the total amount of data involved in the computation, oversegmentation significantly increases the computational cost. Aijazi *et al.* [8] presented a voxel-based segmentation technique using a link-chain method and then classified the segmented objects by employing geometrical and local feature descriptors. On using voxels in reconstructing building models, Truong-Hong *et al.* [9] extracted facade features based on the underlying vertical and horizontal grid voxels of the feature boundaries using a grid clustering technique. The geometric buildings were modeled using the flying voxel method.

In [10], point-based and object-based features were derived from line and plane segments obtained using the random sample consensus and the minimum description length. Two random forest classifiers were combined with the two types of features and used to classify urban point clouds. In [11], features were extracted from three different entities—points, plane segments, and line segments—in a multiple-entity-based classification system. This method achieved a better overall accuracy than the point-based methods did. The methods reported in [3] and [12] involved labeling objects at the object level by extracting their shape features. The advantages and the limitations of an object-based approach in remote sensing image classification were evaluated in [13].

B. Classifiers

In addition to defining discriminative feature descriptors, it is essential to design proper classifiers to recognize the objects in a cluttered scene. Many supervised statistical classifiers have been developed for classifying LiDAR point clouds. For example, for urban mapping, Mallet [14] used a point-based multiclass support vector machine (SVM) to classify full-waveform LiDAR point clouds. A supervised classification method using locally extracted features from a LiDAR point cloud was presented to identify on-ground objects [2]. However, both approaches classified each point independently without considering the labels of neighboring points. Contextual information [15] from neighbors was used to improve the final classification. Each 3-D point in an irregularly distributed point cloud was assigned to a semantic object class. This can be regarded as a tradeoff between objects and pixel orientations. Based on an SVM and evolutionary majority voting, García-Gutiérrez *et al.* [16] developed a contextual classifier, which is called SVM-EMV, to model land use and land cover maps from high-resolution LiDAR data using imagery data fusion. By integrating a random forest classifier into a CRF framework, Niemeyer [17] *et al.* developed a context-based CRF classifier for urban LiDAR point clouds. Based on the construction of a minimum spanning forest from region markers, Tarabalka *et al.* [18] used spatial information described by a Markov random field to update the classification results produced by a probabilistic SVM. Negri *et al.* [19] developed a contextual classification method that consisted of locally adapting the classification hyperplane. They used contextual information to displace the separation hyperplane obtained by the traditional SVM. In [3], neighboring nearby points were clustered hierarchically to form a set of potential object locations, a graph-cut algorithm was used to segment the points surrounding those locations into foreground and background sets, and a shape and a contextual feature were constructed for each point cluster. Finally, an SVM classifier was used to classify the objects into semantic groups.

C. Hierarchical Classification

In hierarchical data structures, local neighborhoods have been determined with respect to either their absolute sizes [20] or their scale parameters [21]. In addition to optimal neighborhoods, features on multiple scales may be considered. A multiscale hierarchical framework was used to describe the classification of terrestrial laser scanning (TLS) point clouds of cluttered urban

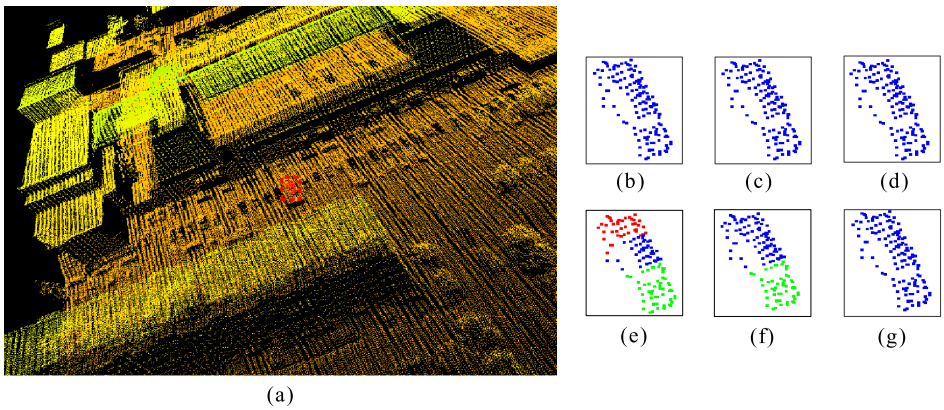


Fig. 2. Hierarchical point-cluster sets of a car generated using method [5] and our method. (a) Point cloud of a scene containing a car, which is shown in red. (b)–(d) Multilevel point clusters generated using method [5] in order from fine to coarse. (e)–(g) Multilevel point clusters generated using our method in order from fine to coarse. The different colors represent different point clusters in (b)–(g).

scenes [5]. In this framework, point clouds were first resampled onto different scales. Then, the resampled data set for each scale was aggregated into several hierarchical point clusters. Another method that has been used in image classification is the superpixel approach [22], [23], which aggregates local neighboring points that have similar colors and texture statistics.

Brodu and Lague [24] classified TLS point clouds using a multiscale local dimensionality feature. Because it combined various scales, the method performed better than a single scale analysis and was robust to missing TLS data. In [25], the authors presented a classification framework that allowed discrete surface analysis at multiple scales. Xiong *et al.* [26] used point cloud statistics and relational information on fine and coarse scales. Xu *et al.* [27] employed three types of entity, namely, single points, plane segments, and segments obtained by mean-shift segmentation, to classify point clouds. In these two methods, different scales were used to determine the context of the point cloud and the shapes of the objects. In [28], a rule-based hierarchical semantic classification scheme that used spectral information, geometry, and topological features was developed. Because multilevel structures are capable of representing the semantic intercorrelations or visual similarities among categories [29], hierarchical dictionary learning models have been used to enhance classification performance [30]–[33]. To avoid losing all information about the spatial layout of a set of features, Zhou *et al.* [34] incorporated a multiresolution representation into a bag-of-features model. They partitioned an image into several images with multiple resolutions and extracted local features from each of the multiresolution images. Then, the representations of the channels at different resolutions were combined, and an SVM was used to reach a final decision.

III. Construction of NEHCS

Discriminative features adapted to object points with different spatial extents are necessary for recognizing objects in ALS point clouds. In [5], the point number of each point cluster at each level changes in an approximately linear manner. Small objects in an ALS point cloud, such as cars, cannot be segmented adequately because they contain few points. For example, Fig. 2(b)–(d) shows that the method described in [5] generates three levels of point clusters for a car, which are illustrated by the red points in Fig. 2(a). We observe that there is only one point cluster at each of the three levels of the car. It is difficult to identify stable and distinct features at different levels.

To solve this problem, we propose a method for generating hierarchical point clusters, i.e., NEHCs, that is based on a natural exponential function and that segments the input point cloud into better hierarchical point-cluster sets. The process for generating the NEHCs includes the following steps.

1) The terrain points and the isolated points are first removed using the method described in [35]. Removing the terrain points helps with the determination of the connectivity of the on-ground objects.
2) Each of the remaining points is connected to its nearest k neighbors to form an undirected graph $G(\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is the set of points, and $\mathbf{E}$ is the set of edges.
3) Because on-ground objects are often close together in cluttered urban scenes, a connected component can contain more than one object. To further break a connected component into smaller pieces so that single objects can be isolated, a moving window algorithm is applied to identify the local maxima in a 2-D raster image. The image represents the heights of the points in the connected component. The raster value is the maximum height of the points in each raster. When the local maxima are found, a graph cut [36] is employed to segment the connected components, and the local maxima are used as seeds. After the graph cut is performed, each connected component is divided into several point clusters.
4) A normalized cut [37] is used to partition a large point cluster into two new clusters when the number of points in the cluster is greater than a predefined threshold δ_m. To ensure that each point cluster contains enough spatial or shape information, we define δ_m at different levels and set $\delta_m = \eta e^x$, where η is a parameter, and x is an integer related with the level number. Thus, the input point cloud is segmented into multilevel point-cluster sets.

The smallest point clusters (i.e., those with $x = 2$) are at the lowest level, namely, the nth level. In the jth $(j < n)$ level, $x = n + 2 - j$. Fig. 3 shows an example of a three-level point cluster hierarchy. Similarly, Fig. 2(e)–(g) shows the car point-cluster sets for the three levels generated using our method. In Fig. 2(e), the points of the car in Fig. 2(a) are segmented into three point clusters. They are segmented into two point

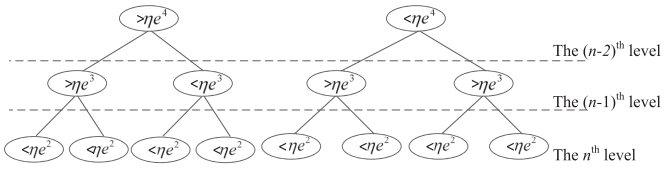


Fig. 3. Illustration of three-level point clusters.

clusters in Fig. 2(f) and into one point cluster in Fig. 2(g). The same-level point clusters shown in Fig. 2(e) and (f) describe the local features of the car. The point cluster in the third level [see Fig. 2(g)] represents the global appearance of the car. Therefore, this approach to partitioning the point cloud can help with the extraction of discriminative features of objects with different sizes from scanned scenes.

IV. Extraction of Point-Cluster-Based Features

Point-based features do not consider the spatial relationships among neighboring points; therefore, they are sensitive to noise and clutter in point clouds. The features of the NEHCs take the spatial relationships among the points and the shape of the point cluster at each level into account. To construct the features of NEHCs, the point-based features are first extracted and then used to construct an SCLDA model. Finally, the features of each NEHC are obtained using the SCLDA model.

Because the point clouds have been split into multiple levels, their point-based features are extracted at each level. The set of the k-nearest neighborhoods of a given point p is defined as the support region of p [5]. Here, $k = 30$, 60, and 90. For each value of k, an 18-dimensional descriptor is obtained from the eigenvalues and the spin image of p, which is based on the support region. Therefore, the resulting 54-dimensional descriptor is considered the feature of p. Next, we focus on the process of deriving the feature of an NEHC.

To efficiently handle the key points that are salient image patches, a bag-of-words (BoW) method [38], [39] is often used to quantize each extracted key point into a visual word and then to represent each image using a histogram of the visual words. Inspired by the ideas behind BoW methods for classifying objects, SCLDA is introduced to derive the features of NEHCs. We know that a set of point clusters generated using the aforementioned method is hierarchical and that the number of points in each point cluster is different. To better express the features of NEHCs using the SCLDA model, we introduce sparse coding to describe the point-cluster-based features. Sparse coding has an obvious advantage in vocabulary extraction and feature representation [40]–[42]; its basic assumption is that the input data can be represented using linear combinations of items in an overcomplete vocabulary. The vocabulary is trained using the point-based features. First, we define each NEHC as a document; all the NEHCs make up a set of documents. The vocabulary generated using sparse coding is considered the vocabulary of the LDA. The point-based feature of each point in an NEHC is treated as a basic unit, and each feature is encoded using sparse coding. The frequency of each word in an NEHC is computed to generate a word frequency vector with length N (where N is the number of words in the vocabulary). The SCLDA model is trained using the point-based features. The process of extracting topics with SCLDA-based hierarchical features is illustrated in Fig. 4. It consists of the following steps.

1) Encode point-based features using sparse coding. A point-based feature $\mathbf{x}$ can be encoded into an N-dimensional vector $\mathbf{u} = [u_1,\ u_2, \ldots,\ u_N]$ by fitting an optimized model with the sparsity constraint

$$\begin{aligned} \min_{\mathbf{V},\mathbf{U}} \quad & \left(\|\mathbf{X} - \mathbf{V}\mathbf{U}\|_2^2 + \lambda|\mathbf{U}|\right) \\ \text{subject to} \quad & \|\mathbf{w}_k\| \le 1,\ \forall k = 1, 2, \ldots, N \end{aligned} \tag{1}$$

where a unit L_2-norm constraint on $\mathbf{w}_k$ is typically applied to extract the discriminative feature, N is the number of words, $\mathbf{V} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N]^T$ is the vocabulary, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_q]^T$ is the set of point-based features, and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_q]$ is the sparse representation of $\mathbf{X}$. The second term in (1) is the L_1 penalty term, which enforces the sparsity of the vector $\mathbf{U}$, and λ is the parameter that controls the sparsity.

2) Extract the vocabulary. In the training phase, (1) is optimized by solving for $\mathbf{U}$ and $\mathbf{V}$. The vocabulary $\mathbf{V}$ can be determined by solving for $\mathbf{U}$ and $\mathbf{V}$ iteratively. Therefore, the feature sign method [43] is used to solve for $\mathbf{U}$ when $\mathbf{V}$ is fixed. The Lagrange dual method [43] is used to solve for $\mathbf{V}$ when $\mathbf{U}$ is fixed. This way, $\mathbf{V}$ is determined iteratively.

3) Create the sparse representation. After $\mathbf{V}$ has been determined, sparse representations are made throughout the vocabulary. The point-based features are encoded using the feature sign method [43].

4) Extract the SCLDA-based feature of the NEHC. A vocabulary $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N]$ that is the same as $\mathbf{V}$ in (1) is set up. The probability of the ith word in the document can be calculated using

$$p(\mathbf{w}_i|\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{k=1}^{r} u_i^k \tag{2}$$

where u_i^k is computed using (1), r is the number of points in the NEHC, $\boldsymbol{\beta}$ is a $K \times N$ matrix, and $\boldsymbol{\theta}$ is a K-dimensional Dirichlet random variable [44], i.e., $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]$, where θ_i represents the probability of the ith topic, and K is the number of topics. A latent topic set $\mathbf{z} = [z_1, z_2, \ldots, z_n]$ is created, and $\boldsymbol{\alpha}$ is defined as a Dirichlet parameter. We build the LDA model [44] as follows:

$$p(\mathbf{w}|\alpha, \beta) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i (\Gamma(\alpha_i))} \int \begin{array}{l} (\prod_{i=1}^{k} \theta_i^{\alpha_i - 1}) \cdot \\ (\prod_{n=1}^{N} \sum_{z_n} p(z_n|\theta) \cdot \\ \quad p(\mathbf{w}_n|z_n, \beta)) d\theta \end{array} \tag{3}$$

In the training phase, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are calculated using the expectation-maximization algorithm because $\mathbf{w}$ is considered a variable. $\boldsymbol{\theta}$ and $\mathbf{z}$ are hidden variables. Finally, the probability of each latent topic in the point cluster is derived. The SCLDA-based feature of the NEHC is defined as $\mathbf{F}_{C_i^j}^{SL}$, and then

$$\mathbf{F}_{C_i^j}^{SL} = [\theta_1, \theta_2, \ldots, \theta_K]. \tag{4}$$

Equation (4) represents the SCLDA-based feature of NEHC C_i^j.

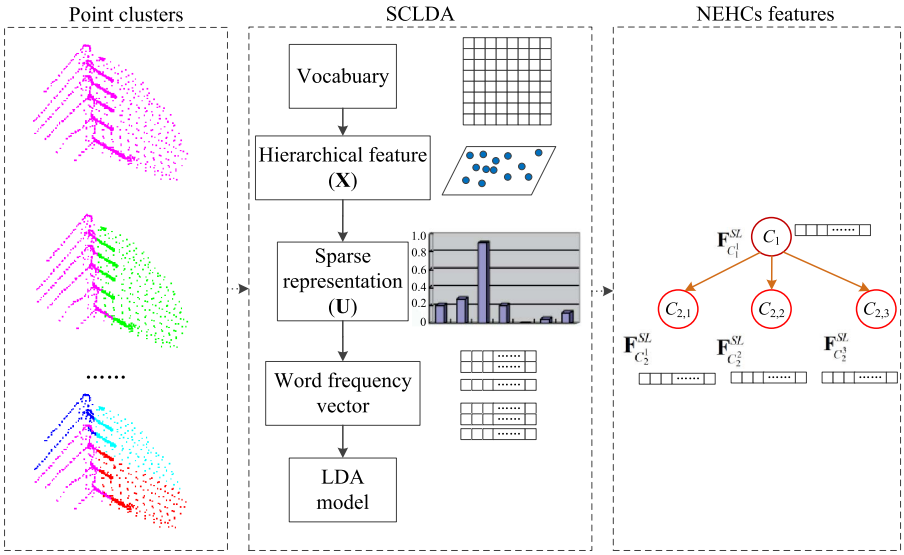


Fig. 4. Process of extracting the SCLDA-based feature of an NEHC.

V. Point Cloud Classification

We use the AdaBoost classifier to classify the point-cluster sets into different categories based on the SCLDA-based features of the NEHCs. The training data are clustered into hierarchical point clusters, and the SCLDA-based features are extracted from the data. Once all the features of the hierarchical point clusters have been derived using the SCLDA-based model, the one-versus-all AdaBoost classifiers are trained. For n point clusters, there are $3n$ AdaBoost classifiers. The SCLDA model and the AdaBoost classifiers are obtained during the training process. They are then used to classify the unlabeled point clouds. The unknown point-cluster sets are labeled according to their hereditability from coarse to fine levels. Following [23], the probability of assigning a label l_i to a specific cluster is mathematically expressed as

$$P_{\text{num}}\left(l_i, \mathbf{F}^{SL}\right) = \frac{\exp\left(H_{\text{num}}(l_i, \mathbf{F}^{SL})\right)}{\sum_i \exp\left(H_{\text{num}}(l_i, \mathbf{F}^{SL})\right)} \tag{5}$$

where $\mathbf{F}^{SL}$ is the SCLDA-based feature of each point cluster, num is an integer $(1 \leq \text{num} \leq n)$, and $H_{\text{num}}(l_i, \mathbf{F}^{SL})$ is the output of the AdaBoost classifier for l_i.

The point-cluster sets are labeled according to their inheritance from coarse to fine levels. During the training process, the training data are manually labeled, and each cluster only contains one specific object category. During the generalization process, the finest point cluster only includes one object or one part of an object, and the point clusters in each of other levels may contain more than one object. Therefore, we only label point-cluster sets at the fine level. A point cluster and the point clusters in its coarse level contain different SCLDA-based features; therefore, unknown point-cluster sets are labeled using the joint probabilities assigned to multiple point-cluster sets. The point clusters at each level inherit the recognition result of the previous level. As shown in Fig. 5, the probability of labeling a cluster C_i in a point-cluster set on the ith level with label l_i is P^i, and the probability of labeling a cluster C_{i+1} in a point-cluster set on the $(i+1)$th with label l_i level is P^{i+1}. Eventually, the classification result of cluster C_{i+1}

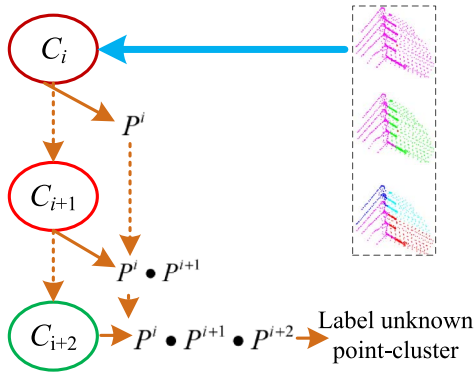


Fig. 5. Labeling unknown point clusters using inheritance.

is $P^i \times P^i + 1$, which is the inherited probability of C_i. In the same way, the probability $P^i \times P^{i+1} \times P^{i+2}$ of the fine point cluster C_{i+2} in each category can be obtained. The final probability of labeling one cluster in a point-cluster set with label l_i can be mathematically expressed as

$$P_n^j(l_i) = \prod_{m=1}^{n} P^{m,\text{num}}(l_i, \mathbf{F}_{SL}) \tag{6}$$

where n denotes the number of levels, P_n^j denotes the probability that the jth point cluster is attributed to category l_i, and $P_n^{m,\text{num}}$ denotes the probability that the mth point-cluster sets on the numth level are attributed to category l_i. Finally, all the point clusters on the finest level are labeled with the label with the highest probability.

VI. Experimental Results

To evaluate the performance of our method, we perform both qualitative and quantitative evaluations on the ALS point clouds of three urban scenes.

A. Experimental Data Sets

We used three data sets from Tianjin, China, which were acquired in August 2010 by a Leica ALS50 system with a

TABLE I
EXPERIMENTAL DATA SETS

	Number of points in the training data			Number of points in the test data		
	Trees	Buildings	Cars	Trees	Buildings	Cars
Scene I	68,802	37,128	5,380	213,990	200,549	7,816
Scene II	39,743	64,952	4,584	73,207	156,186	7,409
Scene III	6,835	13,742	905	33,970	55,459	1,812

TABLE II
MAIN CHARACTERISTICS OF OUR METHOD AND THREE OTHER METHODS

Methods	Feature representation	Vocabulary extraction method	Vocabulary representation	Cluster-based?	Hierarchical point-cluster sets?
Our Method	SCLDA of point-cluster sets	Lagrange dual method	Feature sign method	Yes	Exponential
Method I	LDA of point-cluster sets	k-means method	Vector quantization	Yes	Exponential
Method II	Point-based	No vocabulary	No vocabulary	No	No
Method III	Point-based	No vocabulary	No vocabulary	No	No

mean flying height of 500 m above the ground and a 45° field of view. The average strip overlap was 30%, and the point density in the test areas was approximately 20–30 points/m^2. Large objects, such as buildings and trees, and small objects, such as cars, were present in the three experimental areas. The numbers of test and training points for each type of object after terrain points were filtered and removed are listed in Table I. In Scene I, buildings with different roof shapes, e.g., flat and gable roofs, were surrounded by trees and cars. In Scene II, there were buildings with different heights, dense complex trees, and cars on the roads. In Scene III, cars were moving on the road throughout the scan, and the road was surrounded by trees and buildings.

We used Scene I and Scene II to compare the classification results obtained by our method with those obtained by other methods. The only cars in these two scenes were parked. Cars moved along the road throughout the scan of Scene III. This scene was used to further evaluate the quality of our extracted features in the point cloud classification process, particularly for classifying small objects.

As shown in Table I, the amounts of data in each category were different in each of the three areas. We manually labeled the point clouds of the three areas to enable the evaluation of the classification results.

We implemented the proposed method in C++. The algorithm ran on a computer with a 3.40-GHz Intel Core i7-4770 K processor and 8 GB of random access memory. The system took approximately 25.3 min to learn the SCLDA-based models and the AdaBoost classifiers. It took approximately 23.9, 10.9, and 7.3 min to classify the point clouds of Scene I, Scene II, and Scene III, respectively. During processing, feature extraction and sparse representation took approximately 67% of the running time. However, most of the steps are parallelizable. Therefore, they can be implemented using a parallel scheme to reduce the amount of time required.

B. Comparison With Other Methods

To validate the performance of our method, we compared it with three other methods. The first method (Method I) used a combination of a BoW and LDA of point-cluster sets to classify the unknown data. The vocabulary was extracted using *k*-means, and each point-based feature was represented using vector quantization. The second method (Method II) used point-based features. It used point-based features to classify point clouds directly without aggregating the data into point clusters. Method II did not involve hierarchical structures. The third method (Method III) was the one described in [2]. In this method, each point was associated with a set of defined features derived using geometric, multireturn, and intensity information; and features were selected using JointBoost to evaluate their correlations. On the other hand, our method does not utilize the contextual information, such as intensity, to classify the point clouds. Table II shows the main characteristics of our method and the other three methods.

We selected some points from each scene to form the training data set, which is illustrated in Fig. 6. The complete training data set contained 242 071 points, including 115 380 tree points, 115 822 building points, and 10 869 car points. The numbers of points in each category in Scene I, Scene II, and Scene III are listed in Table I. Considering the diversity of the objects and the completeness of the object points, we selected points on buildings with different roofs, trees with different heights, and parked and moving cars from Scene I, Scene II, and Scene III, respectively.

Precision/recall can be used to represent the quality of the classification [5]. Precision is the fraction of the retrieved instances that are relevant, and recall is the fraction of the relevant instances that are retrieved. High precision means that the algorithm returned substantially more relevant results than irrelevant results, and high recall means that the algorithm returned most of the relevant results. Table III shows the precision/recall of the four methods in the learning stage. As shown in Table III, the precision and recall of the classification results obtained by our method are the highest for the classification of objects in the three specific classes. The car classification precision of our method is much higher than that of the other methods. Therefore, SCLDA-based features describe the training data and distinguish objects better than other features.

The quality of the classifications performed using the methods earlier has been also tested. Table IV shows the classification results of the two scenes in terms of their precision/recall

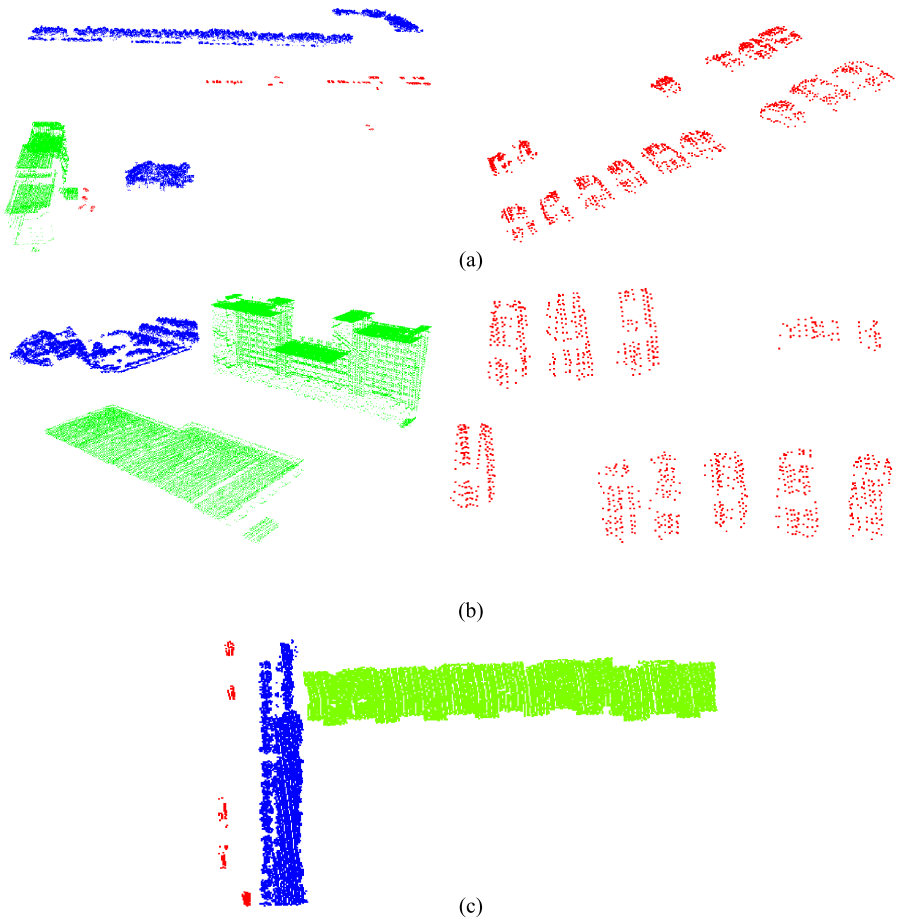


Fig. 6. Training data. Blue points represent trees, green points represent buildings, and red points represent cars. (a) Part of the training dataset obtained from Scene I. (b) Part of the training dataset obtained from Scene II. (c) Part of the training dataset obtained from Scene III.

TABLE III
PRECISION/RECALL AND THE ACCURACY OF DIFFERENT METHODS IN THE LEARNING STAGE

	Scene	Trees (%)	Buildings (%)	Cars (%)	Accuracy (%)
Our method	I	**97.5/98.9**	**98.2/98.9**	**95.8/87.8**	**97.8**
	II	**92.9/98.7**	**99.3/98.5**	**89.7/59.5**	**96.9**
Method I	I	96.9/97.1	95.7/94.5	77.7/82.3	96.6
	II	94.5/97.4	98.8/98.4	72.5/54.6	96.2
Method II	I	89.7/95.6	92.6/84.5	79.1/59.5	90.2
	II	78.3/91.5	92.6/87.9	35.1/25.5	86.6
Method III	I	95.9/97.7	96.1/96.5	85.7/63.5	95.6
	II	91.3/92.6	96.6/96.6	63.2/55.5	93.4

and their accuracies. Except for a few points on the corners of buildings and on cars that were difficult to distinguish, most of the points were correctly recognized by our method. As shown in Table IV, the precision and recall of our method with respect to cars are higher than those of the other three methods. In addition, the final classification accuracy of our method was higher than that of the other three methods.

Figs. 7 and 8 visually show the classification results. Compared with other classifiers, our method was more accurate; it classified most of the points into the correct categories, including the cars in dashed boxes in Fig. 7 and in dashed box in Fig. 8. The classification result of Method II was the worst, which means that using clusters as basic units helps improve the classification results. Method I performed better at classifying trees and buildings than at classifying cars; however, it did not effectively recognize the trees, buildings, and cars in the cluttered point clouds. Although Method III classified the objects well, the results for cars need further improvement. In our method, vocabulary training and representing features using sparse coding result in accurate expressions of the characteristics of the objects. Furthermore, sparse coding combined with LDA to represent point-cluster features leads to high-precision classification results.

C. Classification of Scene III

We classified the objects in Scene III using our method. The result is shown in Fig. 9(b). Compared with Fig. 9(a), most of

TABLE IV
PRECISION/RECALL AND ACCURACY OF THE CLASSIFICATION RESULTS

Scene I	Trees (%)	Buildings (%)	Cars (%)	Accuracy (%)
Our method	**95.**7/96.2	95.9/**95.9**	**80.8**/**67.9**	**95.8**
Method I	94.0/95.4	95.0/94.3	79.1/60.8	94.5
Method II	85.7/92.9	92.0/83.8	56.9/54.7	87.9
Method III	89.7/**98.1**	**97.9**/89.1	65.2/46.6	92.9
Scene II	Trees (%)	Buildings (%)	Cars (%)	Accuracy (%)
Our method	**94.7**/**94.5**	**98.1**/**97.7**	**53.9**/**60.5**	**95.5**
Method I	90.3/93.9	97.6/96.5	49.4/42.0	94.1
Method II	73.9/91.2	93.6/88.2	29.5/25.4	87.2
Method III	86.8/91.2	96.8/95.5	44.1/34.8	92.2

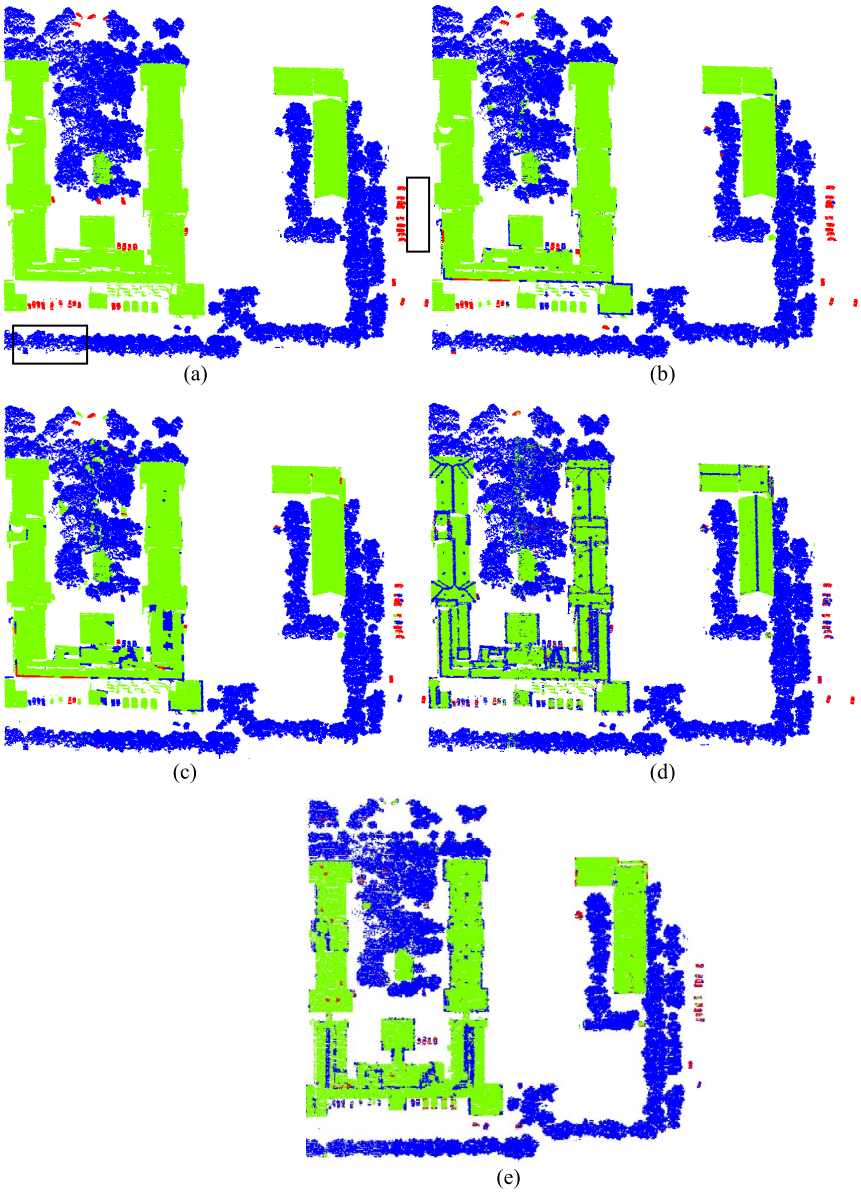


Fig. 7. Classification results obtained using the different methods for part of Scene I. (a) Ground truth. (b) Classification results obtained by our method. (c) Classification results obtained by Method I. (d) Classification results obtained by Method II. (e) Classification results obtained by Method III. The points on trees, buildings, and cars are colored blue, green, and red, respectively. Our method correctly classifies most of the points and outperforms the other three methods.

the points in Scene III are correctly recognized in Fig. 9(b), particularly those points on moving cars. As shown in Table V, our method's classification results for Scene III were very precise.

D. Sensitivity of the Parameters

Here, we analyze the impact of the parameters, including the point-cluster generation threshold, the numbers of topics and words, and the sparse coding parameters, on the classification

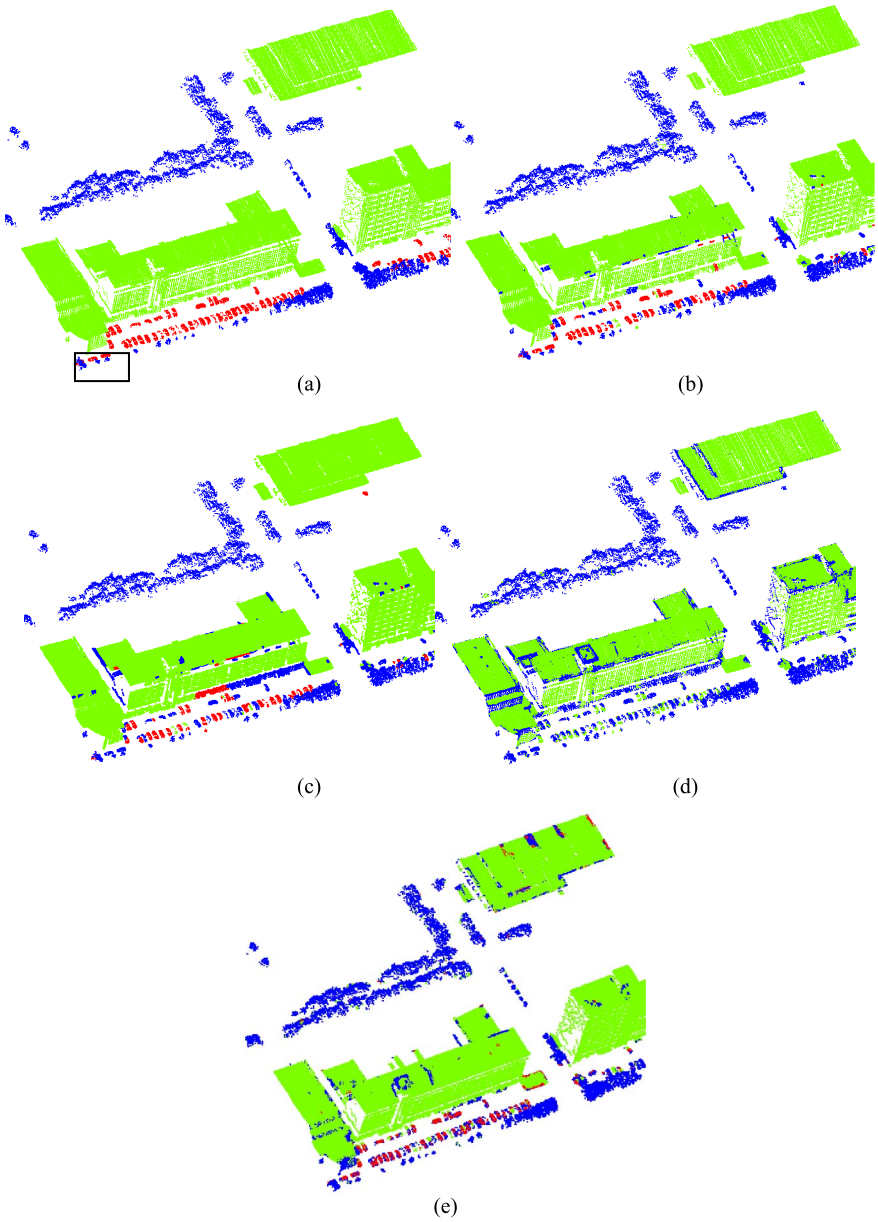


Fig. 8. Classification results obtained using different methods for part of Scene II. (a) Ground truth. (b) Classification results obtained by our method. (c) Classification results obtained by Method I. (d) Classification results obtained by Method II. (e) Classification results obtained by Method III. The points on trees, buildings, and cars are colored blue, green, and red, respectively. Our method correctly classifies most of the points and outperforms the other three methods.

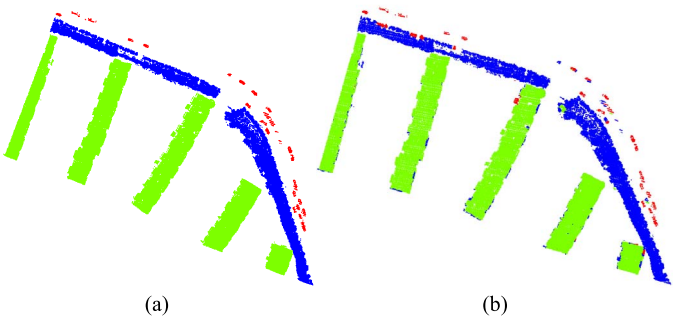


Fig. 9. Classification of Scene III, which includes a point cloud for moving cars. (a) Ground truth. (b) Classification results obtained by our method. The points on trees, buildings, and cars are colored blue, green, and red, respectively.

results. The F_1 measure expressed by the following is used to represent the quality of the classifications of Scene I and Scene II:

$$F_1 = \frac{2(\text{recall} \times \text{precision})}{\text{recall} + \text{precision}}. \tag{7}$$

1) Value of the Point-Cluster Set Generation Threshold: We test the effect of different values of the point-cluster set generation threshold δ_m, which is controlled by η in the formula $\delta_m = \eta e^x$, on F_1. Let η be equal to 8, 10, 12, and 14 when there are 512 words and ten topics and when $\lambda = 0.15$. The results are shown in Fig. 10. For Scene I and Scene II, our method almost maintains the stability of the recognition of the three categories, and it results in a better value of F_1 for the recognition of the three categories when $\eta = 10$.

TABLE V
PRECISION/RECALL AND THE ACCURACY OF THE CLASSIFICATION RESULTS FOR SCENE III

Scene III	Trees (%)	Buildings (%)	Cars (%)	Accuracy (%)
Our method	87.4/97.7	98.8/91.3	70.7/77.1	93.4

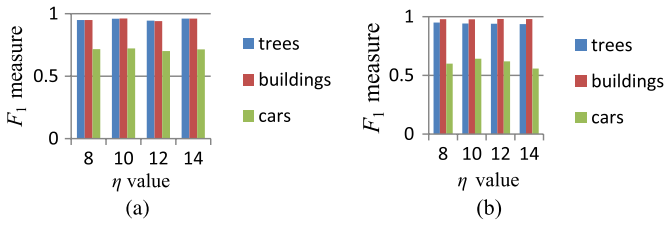


Fig. 10. Effect of different values of the point-cluster set generation threshold δ_m on the classification results. (a) Effect of different values of η on the classification results for Scene I. (b) Effect of different values of η on the classification results for Scene II.

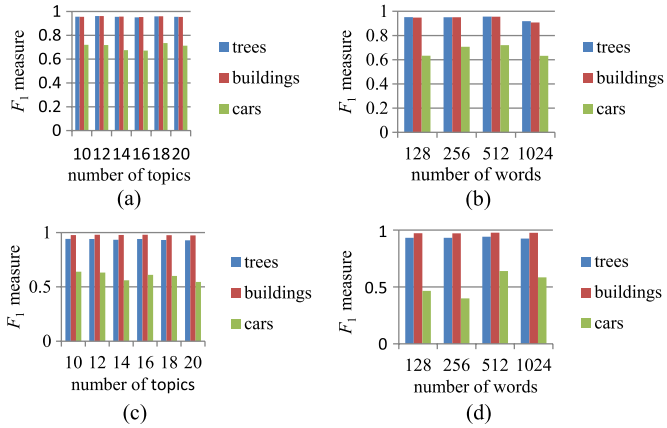


Fig. 11. Effect of different numbers of words and topics on the classification results. (a) Effect of different numbers of topics on the classification results for Scene I. (b) Effect of different numbers of words on the classification results for Scene I. (c) Effect of different numbers of topics on the classification results for Scene II. (d) Effect of different numbers of words on the classification results for Scene II.

2) Numbers of Topics and Words: Fig. 11 shows the quality of the classifications of Scene I and Scene II measured using F_1 for different numbers of topics and words. When $\eta = 10$, the number of words is 512, and the sparsity is $\lambda = 0.15$, we select 10, 12, 14, 16, 18, and 20 topics to show the effect of different numbers of topics on the quality of the classification [see Fig. 10(a) and (c)]. For the same values of η and λ, we also select 128, 256, 512, and 1024 words and ten topics to analyze the impact of the number of words on the quality of the classification [see Fig. 10(b) and (d)].

Fig. 11(a)–(d) shows that the numbers of words and topics have little effect on the quality of the classification of the three categories in Scene I and Scene II. Most of the differences in the values of F_1 for the three categories are less than 9% (see Table V), which means that our method performs well and stably when there are few words and topics. The number of topics influences the recognition of trees and buildings less; for these two categories, the standard deviations of F_1 are 0.0036 and 0.0032, respectively, for Scene I, and 0.0056 and 0.0025, respectively, for Scene II. The number of words also slightly influences the quality of the classification of trees and buildings because the features on the surfaces of trees and buildings are simple

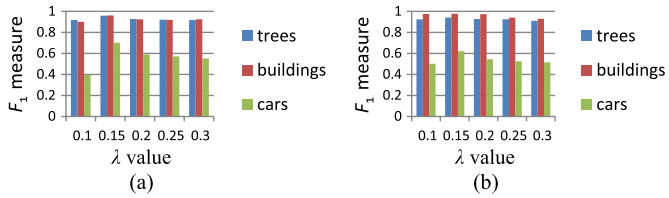


Fig. 12. Effect of different values of λ on the classification results. (a) Effects of different values of λ on the classification results for Scene I. (b) Effect of different values of λ on the classification results for Scene II.

and obvious. Selecting a few words and topics is sufficient to describe these discriminative features. As shown in Fig. 11, the values of F_1 for buildings and trees are no less than 0.9. When the numbers of topics and words increase, the value of F_1 for cars fluctuates. The value of F_1 for the cars in Scene I is greatest when the numbers of topics and words are 18 [see Fig. 11(a)] and 512 [see Fig. 11(b)], respectively. For Scene II, the value of F_1 for the cars is highest when the number of topics is 10 [see Fig. 11(c)] and the number of words is 512 [see Fig. 11(d)]. Buildings and trees have the highest values of F_1 when the number of topics is 10 [see Fig. 11(c)] and the number of words is 512 [see Fig. 11(d)]. The values of F_1 for buildings and trees remain stable when the numbers of topics and words change.

Table VI lists the standard deviations of the values of F_1 for the three categories for different numbers of topics and words. As shown in this table, the values of F_1 for trees and buildings are no more than 0.56%, which demonstrates the stability of the classification results for buildings and trees. The value of F_1 for cars changes when the numbers of topics and words change because there are fewer cars in the training sample to extract the vocabulary. However, our method still provides good classification results.

3) Sparse Coding Parameters: There is one parameter, i.e., λ, in (1). When there are 512 words and 10 topics and when $\eta = 10$, we test different values of λ, i.e., $\lambda = 0.1, 0.15, 0.2, 0.25$, and 0.3. Fig. 12 shows the impacts of different values of λ on the classification results. Using $\lambda = 0.15$ provides good results for the two scenes.

E. Error Analysis

We analyze the classification results of Scene I and Scene II when there are 512 words and 10 topics and when $\lambda = 0.15$ and $\eta = 12$. Tables VII and VIII list the confusion matrices of the two scenes. The points on the prominent eaves are often incorrectly classified as car or tree points [see dashed box in Fig. 13(a)]. Due to the inhomogeneity of the input point cloud, the points on some cars are distributed linearly, as shown in the dashed box in Fig. 13(a); these are misclassified as tree points. Some car points are scattered so that they are often recognized incorrectly, as the points in the dashed box in Fig. 13(c) are. Some tree points are misclassified as points on building roofs because some local point clusters are nearly flat [see dashed box 1 in Fig. 13(d)]; however, the generated point cluster for a single tree crown [see dashed box 2 in Fig. 13(d)] may be incorrectly classified as a car because its shape is similar to that of a point cluster of a car.

VII. CONCLUSION

This paper has presented a method for extracting object shape features for ALS point cloud classification. In this method, the

TABLE VI
STANDARD DEVIATIONS OF F_1 FOR DIFFERENT NUMBERS OF TOPICS AND WORDS

	Scene I		Scene II	
	10-20 topics, 512 words	10 topics, 128-1024 words	10-20 topics, 512 words	10 topics, 128-1024 words
Trees	0.0036	0.0175	0.0056	0.0067
Buildings	0.0032	0.0220	0.0025	0.0029
Cars	0.0260	0.0470	0.0381	0.0712

TABLE VII
CONFUSION MATRIX FOR THE RESULTS OF CLASSIFYING SCENE I

	Trees	Buildings	Cars	Recall
Overall accuracy: 94.28%				
Trees	203,017	10,449	524	0.9487
Buildings	11,158	187,967	1424	0.9373
Cars	1967	583	5266	0.6737
Precision	0.9393	0.9446	0.7300	

TABLE VIII
CONFUSION MATRIX FOR THE RESULTS OF CLASSIFYING SCENE II

	Trees	Buildings	Cars	Recall
Overall accuracy: 96.48%				
Trees	70,304	1762	1141	0.9603
Buildings	3200	152,503	483	0.9764
Cars	2902	454	4053	0.5470
Precision	0.9201	0.9857	0.7139	0.9649

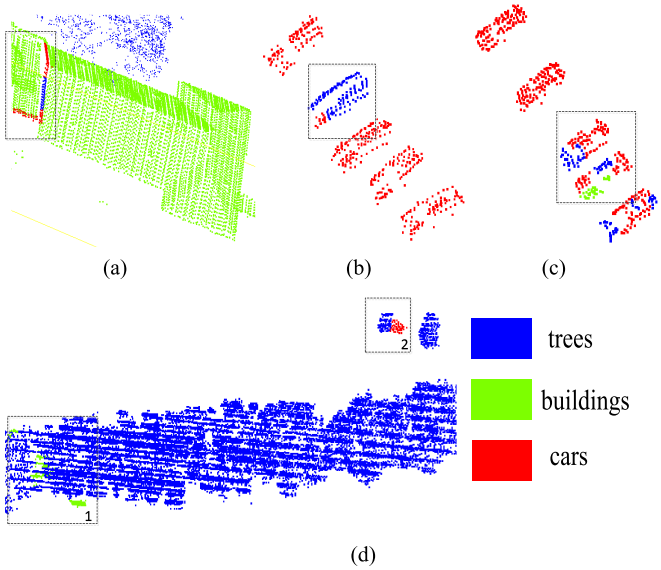


Fig. 13. Typical misclassification errors. (a) Points on building edges are misclassified as points on trees and cars. (b) Points on cars are misclassified as points on trees. (c) Points on cars are misclassified as points on trees and buildings. (d) Points on trees are misclassified as points on buildings and cars.

NEHCs divide the ALS point cloud into multiple levels, and in each level, the data are aggregated into several point clusters with different sizes. Splitting the point cloud guarantees that each point cluster includes the points necessary for extracting the shape features of small objects. SCLDA is exploited to extract and encode multilevel discriminative features. Then, each point cluster is labeled using the joint probability of assigning a particular label over the multiple point-cluster sets. The SCLDA is flexible enough to extend the classifiers to multifeature channels, which further enhances the classification accuracy. The experimental results show that our method takes advantage of correlations between the multilevel point clusters to effectively capture discriminative information via SCLDA and is capable of performing complex point cloud classification to obtain very precise results, particularly for small objects.

The hierarchy of a point cloud can be built using the semantic correlations or visual similarities between categories. In the future, we will incorporate hierarchical learning and multiple features [45], [46] into this framework to further improve the precision of the classification.

ACKNOWLEDGMENT

The authors would like to thank the editors and anonymous reviewers for their valuable comments.

REFERENCES

[1] C. Mallet and F. Bretar, "Full-waveform topographic LiDAR: State-of-the-art," *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 1, pp. 1–16, Jan. 2009.
[2] B. Guo, X. Huang, F. Zhang, and G. Sohn, "Classification of airborne laser scanning data using JointBoost," *ISPRS J. Photogramm. Remote Sens.*, vol. 100, pp. 71–83, Feb. 2015.
[3] 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.*, Kyoto, Japan, 2009, pp. 2154–2161.
[4] J. Niemeyer, C. Mallet, F. Rottensteiner, and U. Soergel, "Conditional random field for the classification of LiDAR point clouds," in *Proc. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, 38 (Part 4/W19) (on CD-ROM), Aug. 2012, pp. 1–6.
[5] 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.
[6] E. H. Lim and D. Suter, "Multi-scale conditional random fields for over-segmented irregular 3D point clouds classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Anchorage, AK, USA, 2008, pp. 1–7.
[7] E. H. Lim and D. Suter, "3D terrestrial LiDAR classifications with super-voxels and multi-scale conditional random fields," *Comput.-Aided Design*, vol. 41, no. 10, pp. 701–710, Oct. 2009.
[8] A. K. Aijazi, P. Checchin, and L. Trassoudaine, "Segmentation based classification of 3D urban point clouds: A super-voxel based approach with evaluation," *Remote Sens.*, vol. 5, no. 4, pp. 1624–1650, Mar. 2013.
[9] L. Truong-Hong, D. F. Laefer, T. Hinks, and H. Carr, "Combining an angle criterion with voxelization and the flying voxel method in reconstructing building models from LiDAR data," *Comput.-Aided Civil Infrastructure Eng.*, vol. 28, no. 2, pp. 112–129, Feb. 2013.
[10] H. B. Kim and G. Sohn, "Random forests-based multiple classifier system for power-line scene classification," in *Proc. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, 38 (Part 5/W12) (on CD-ROM), 2011, pp. 253–258.
[11] S. Xu, G. Vosselman, and S. Elberink, "Multiple-entity based classification of airborne laser scanning data in urban areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 88, pp. 1–15, Feb. 2014.

[12] A. Patterson, IV, P. Mordohai, and K. Daniilidis, "Object detection from large-scale 3D datasets using bottom-up and top-down descriptors," in *Proc. Eur. Conf. Comput. Vis.*, Marseille, France, 2008, pp. 553–566.
[13] D. Liu and F. Xia, "Assessing object-based classification: Advantages and limitations," *Remote Sens. Lett.*, vol. 1, no. 4, pp. 187–194, Dec. 2010.
[14] C. Mallet, "Analysis of full-waveform LiDAR data for urban area mapping," Ph.D. thesis, Inst. Géographique Nat., Télécom ParisTech, Paris, France, 2010.
[15] F. J. Cortijo and N. Blanca, "Improving classical contextual classifications," *Int. J. Remote Sens.*, vol. 19, no. 8, pp. 1591–1613, Aug. 1998.
[16] J. García-Gutiérrez, D. Mateos-García, M. Garcia, and J. Riquelme-Santos, "An evolutionary-weighted majority voting and support vector machines applied to contextual classification of LiDAR and imagery data fusion," *Neurocomputing*, vol. 163, no. 2, pp. 17–24, Sep. 2015.
[17] J. Niemeyer, F. Rottensteiner, and U. Soergel, "Contextual classification of LiDAR data and building object detection in urban areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 152–165, Jan. 2014.
[18] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1267–1279, Oct. 2010.
[19] R. G. Negri, L. V. Dutra, and S. J. S. Sant'Anna, "An innovative support vector machine based method for contextual image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 241–248, Jan. 2014.
[20] M. Weinmann, B. Jutzi, and C. Mallet, "Feature relevance assessment for the semantic interpretation of 3D point cloud data," in *Proc. ISPRS Ann. Photogramm., Remote Sens., Spatial Inf. Sci.*, 2013, vol. II-5/W2, pp. 313–318.
[21] M. Weinmann, S. Urban, S. Hinz, B. Jutzi, and C. Mallet, "Distinctive 2D and 3D features for automated large-scale scene analysis in urban areas," *Comput Graph.*, vol. 49, pp. 47–57, Jun. 2015.
[22] B. C. Russell, W. T. Freeman, A. A. Efros, J. Sivic, and A. Zisserman, "Using multiple segmentations to discover objects and their extent in image collections," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, 2006, pp. 1605–1614.
[23] J. Xiao and L. Quan, "Multiple view semantic segmentation for street view images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Kyoto, Japan, 2009, pp. 686–693.
[24] N. Brodu and D. Lague, "3D terrestrial LiDAR data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology," *ISPRS J. Photogramm. Remote Sens.*, vol. 68, pp. 121–134, Mar. 2012.
[25] M. Pauly, R. Keiser, and M. Gross, "Multi-scale feature extraction on point-sampled models," *Comput. Graph. Forum*, vol. 22, no. 3, pp. 281–289, Nov. 2003.
[26] X. Xiong, D. Munoz, J. A. Bagnell, and M. Hebert, "3-D scene analysis via sequenced predictions over points and regions," in *Proc. IEEE Int. Conf. Robot. Autom.*, Shanghai, China, 2011, pp. 2609–2616.
[27] S. Xu, S. Oude Elberink, and G. Vosselman, "Entities and features for classification of airborne laser scanning data in urban area," in *Proc. ISPRS Ann. Photogramm., Remote Sens., Spatial Inf. Sci.*, 2012, vol. I-4, pp. 257–262.
[28] J. Rau, J. Jhan, and Y. Hsu, "Oblique aerial images for land cover and pointcloud classification in an urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1304–1319, Mar. 2015.
[29] L. Shen, G. Sun, Q. Huang, Z. Lin, and E. Wu, "Multi-level discriminative dictionary learning with application to large scale image classification," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3109–3123, Oct. 2015.
[30] S. Bengio, J. Weston, and D. Grangier, "Label embedding trees for large multiclass task," in *Proc. NIPS*, 2010, pp. 163–171.
[31] L.-J. Li, C. Wang, Y. Lim, D. M. Blei, and L. Fei-Fei, "Building and using a semantivisual image hierarchy," in *Proc. IEEE CVPR*, Jun. 2010, pp. 3336–3343.
[32] T. Gao and D. Koller, "Discriminative learning of relaxed hierarchy for large-scale visual recognition," in *Proc. IEEE ICCV*, Nov. 2011, pp. 2072–2079.
[33] R. Salakhutdinov, A. Torralba, and J. Tenenbaum, "Learning to share visual appearance for multiclass object detection," in *Proc. IEEE CVPR*, Jun. 2011, pp. 1481–1488.
[34] L. Zhou, Z. Zhou, and D. Hu, "Scene classification using a multi-resolution bag-of-features model," *Pattern Recognit.*, vol. 46, pp. 424–433, Jan. 2013.
[35] D. Chen, L. Zhang, Z. Wang, and H. Deng, "A mathematical morphology-based multi-level filter of LiDAR data for generating DTMs," *Sci. Chin. Inf. Sci.*, vol. 56, no. 10, pp. 1–14, Oct. 2013.
[36] 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.
[37] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
[38] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Proc. Eur. Conf. Comput. Vis.*, Graz, Austria, 2006, pp. 490–503.
[39] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in *Proc. Int. Workshop Multimedia Inf. Retrieval*, New York, NY, USA, 2007, pp. 197–206.
[40] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE CVPR*, Jun. 2009, pp. 1794–1801.
[41] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, Apr. 2012.
[42] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based Fisher discrimination dictionary learning for image classification," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, Sep. 2014.
[43] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 801–808.
[44] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
[45] L. Zhang *et al.*, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognit.*, vol. 48, no. 10, pp. 3102–3112, Oct. 2015.
[46] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.

Zhenxin Zhang is currently working toward the Ph.D. degree in geoinformatics in the School of Geography, Beijing Normal University, Beijing, China.

His research interests include light detection and ranging data processing, quality analysis of geographic information systems, and algorithm development.

Liqiang Zhang received the Ph.D. degree in geoinformatics from the Institute of Remote Sensing Applications, Chinese Academy of Science, Beijing, China, in 2004.

He is currently a Professor with the School of Geography, Beijing Normal University, Beijing. His research interests include remote sensing image processing, 3-D urban reconstruction, and spatial object recognition.

Xiaohua Tong received the Ph.D. degree from Tongji University, Shanghai, China, in 1999.

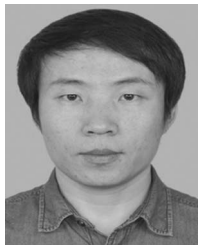
Between 2001 and 2003, he was a Postdoctoral Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China. In 2006, he was a Research Fellow with The Hong Kong Polytechnic University, Hong Kong. Between 2008 and 2009, he was a Visiting Scholar with the University of California, Santa Barbara, CA, USA. He is currently a Chang Jiang Scholar Chair Professor appointed by the Ministry of Education of China with the School of Surveying and Geo-Informatics, Tongji University. In 2013, he received National Natural Science Funds for Distinguished Young Scholar. He has authored over 40 publications in international journals. His current research interests include remote sensing, geographical information systems, trust in spatial data, and image processing for high-resolution and hyperspectral images.

Prof. Tong serves as the Vice Chair of the Commission on Spatial Data Quality of the International Cartographical Association and the Cochair of the International Society for Photogrammetry and Remote Sensing working group (WG II/4) on Spatial Statistics and Uncertainty Modeling. He was a recipient of the State Natural Science Award (second place) from the State Council of China in 2007.

P. Takis Mathiopoulos (SM'94) received the Ph.D. degree in digital communications from the University of Ottawa, Ottawa, ON, Canada, in 1989.

From 1982–1986, he was with Raytheon Canada Ltd., working in the areas of air navigational and satellite communications. In 1988, he joined the Department of Electrical and Computer Engineering (ECE), University of British Columbia (UBC), Canada, where he was a faculty member until 2003, holding the rank of Professor from 2000–2003. From 2000–2014 he was with the Institute for Space Applications and Remote Sensing (ISARS), National Observatory of Athens (NOA), first as Director and then as Director of Research and where he established the Wireless Communications Research Group. As ISARS' Director (2000–2004), he led the Institute to significant expansion R&D growth and international scientific recognition. For these achievements, ISARS has been selected as a national Centre of Excellence for the years 2005–2008. Since 2014, he has been an Adjunct Researcher at the Institute of Astronomy, Astrophysics, Space Applications and Remote Sensing (IAASARS) of NOA. Since 2003 he also taught part-time at the Department of Informatics and Telecommunications, University of Athens, where since 2014 he has been a Professor of Digital Communications. From 2008–2013 he was appointed as a Guest Professor at the Southwest Jiaotong University, China. He has been also appointed by the Government of PR of China as a Senior Foreign Expert at the School of Information Engineering, Yangzhou University (2014–2015) and by Keio University as a Visiting Professor in the Department of Information and Computer Science (2015–2016) under the Top Global University Project of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Government of Japan. For the last 25 years he has been conducting research mainly on the physical layer of digital communication systems for terrestrial and satellite applications, including digital communications over fading and interference environments. He coauthored a paper in GLOBECOM'89 establishing for the first time in the open technical literature the link between MLSE and multiple (or multisymbol) differential detection for the AWGN and fading channels. He is also interested in channel characterization and measurements, modulation and coding techniques, synchronization, SIMO/MIMO, UWB, OFDM, software/cognitive radios, and green communications. In addition, since 2010, he has been actively involved with research activities in the fields of remote sensing, LiDAR systems, and photogrammetry. In these areas, he has coauthored more than 100 journal papers, mainly published in various IEEE and IET journals, 4 book chapters, and more than 120 conference papers. He has been a Principal Investigator for more than 40 research grants and has supervised the thesis of 11 PhD and 23 Master students.

Dr. Mathiopoulos has been or currently serves on the editorial board of several archival journals, including the IET Communications, and the IEEE TRANSACTIONS ON COMMUNICATIONS (1993–2005). He has regularly acted as a consultant for various governmental and private organizations. Since 1993, he has served on a regular basis as a scientific advisor and a technical expert for the European Commission (EC). In addition, from 2001–2014 he has served as a Greek representative to high-level committees in the European Commission (EC) and the European Space Agency (ESA). He has been a member of the Technical Program Committee (TPC) of more than 70 international IEEE conferences, as well as TPC Vice Chair for the 2006-S IEEE VTC and 2008-F IEEE VTC as well as Co-Chair of FITCE2011. He has delivered numerous invited presentations, including plenary and keynote lectures, and has taught many short courses all over the world. As a faculty member at the ECE of UBC, he was elected ASI Fellow and a Killam Research Fellow. He was a corecipient of two best paper awards for papers published in the 2nd International Symposium on Communication, Control, and Signal Processing (2008) and 3rd International Conference on Advances in Satellite and Space Communications (2011).

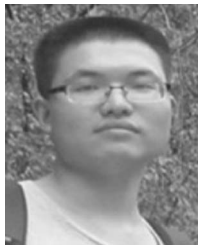


Bo Guo received the Ph.D. degree from Wuhan University, Wuhan, China, in 2014.

He is currently a Postdoctoral Researcher with the Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen, China. His research interests include application of laser scanning point cloud on photogrammetry and computer vision.

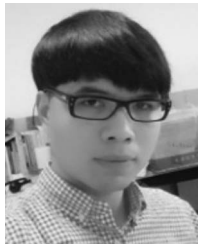
Xianfeng Huang received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2006.

In 2011, he was a Visitor Researcher with Microsoft Research Asia, Beijing, China. From 2012 to 2013, he was a Senior Researcher with Singapore ETH Center. He is currently a Professor with the State Key Laboratory of Information Engineering in Mapping, Surveying and Remote Sensing, Wuhan University. His research interests include laser scanning data processing and photogrammetry.



Zhen Wang received the Ph.D. degree in geographical information systems from Beijing Normal University, Beijing, China.

He is currently with the Department of Geoinformatics, China University of Geosciences, Beijing. His research interests are the application of light detection and ranging in classification of the ground objects and 3-D urban modeling.



Yuebin Wang is currently working toward the Ph.D. degree in the School of Geography, Beijing Normal University, Beijing, China.

His research interests are remote sensing imagery processing and 3-D urban modeling.