A Multiscale and Hierarchical Feature Extraction Method for Terrestrial Laser Scanning Point Cloud Classification

Zhen Wang, Liqiang Zhang, Tian Fang, *Member, IEEE*, P. Takis Mathiopoulos, *Senior Member, IEEE*, Xiaohua Tong, Huamin Qu, *Member, IEEE*, Zhiqiang Xiao, Fang Li, and Dong Chen

***Abstract*—The effective extraction of shape features is an important requirement for the accurate and efficient classification of terrestrial laser scanning (TLS) point clouds. However, the challenge of how to obtain robust and discriminative features from noisy and varying density TLS point clouds remains. This paper introduces a novel multiscale and hierarchical framework, which describes the classification of TLS point clouds of cluttered urban scenes. In this framework, we propose multiscale and hierarchical point clusters (MHPCs). In MHPCs, point clouds are first resampled into different scales. Then, the resampled data set of each scale is aggregated into several hierarchical point clusters, where the point cloud of all scales in each level is termed a point-cluster set. This representation not only accounts for the multiscale properties of point clouds but also well captures their hierarchical structures. Based on the MHPCs, novel features of point clusters are constructed by employing the latent Dirichlet allocation (LDA). An LDA model is trained according to a training set. The LDA model then extracts a set of latent topics, i.e., a feature of topics, for a point cluster. Finally, to apply the introduced features for point-cluster classification, we train an AdaBoost classifier in each point-cluster set and obtain the corresponding classifiers to separate the TLS point clouds with varying point density and data missing into semantic regions. Compared with other methods, our features achieve the best classification results for buildings, trees, people, and cars from TLS point clouds, particularly for small and moving objects, such as people and cars.**

***Index Terms*—AdaBoost, latent Dirichlet allocation (LDA), multiscale and hierarchical point clusters (MHPCs), object classification.**

Manuscript received November 15, 2013; revised April 2, 2014 and July 29, 2014; accepted September 4, 2014. This work was supported by the National Natural Science Foundation of China under Grant 41371324.

Z. Wang, L. Zhang, Z. Xiao, and F. Li are with the State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, Beijing 100875, China (e-mail: comige@gmail.com; zhanglq@bnu.edu.cn; lifang724@126.com).

T. Fang and H. Qu are with the Hong Kong University of Science and Technology, Kowloon, Hong Kong (e-mail: tianft@cse.ust.hk; huamin@cse.ust.hk).

P. T. Mathiopoulos is with the Institute for Astronomy Astrophysics Space Applications and Remote Sensing, National Observatory of Athens, 14536 Athens, Greece, and also with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, 15784 Athens, Greece (e-mail: mathio@hol.gr).

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

D. Chen is with the College of Civil Engineering, Nanjing Forestry University, Nanjing 210037, China (e-mail: chendong@njfu.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.2014.2359951

I. Introduction

THE classification of ground objects is one of the fundamental problems in scene understanding, which has a wide application in areas such as urban planning, virtual tourism, and entertainment. Terrestrial laser scanning (TLS) has been used to obtain 3-D spatial information about objects as point clouds. The classification of point clouds is a key step for using the information effectively. However, the point density in a scene is not uniform because of the varying distance between the objects and the scanner. Even worse, severe occlusion between objects yields incomplete point clouds and cluttered background in urban scenes. Clearly, the extraction of a discriminative feature from the noisy and cluttered point clouds of urban scenes still remains an open and challenging research area, which is also the subject of this paper.

Since the problem of accurate classification of point clouds has been extensively studied for many years, many different classifiers have been proposed. Many different classifiers, such as support vector machines [1], AdaBoost [2], random forests [3], and conditional random fields [4] are used to classify the point clouds. However, as noted in [2], the accuracy of the classifiers is similar and is dependent on the data. Except the classifiers, features also affect the classification performance. In this paper, instead of improving the classifier, we focus on obtaining discriminative shape features to describe the cluster of the point cloud, which is the very first step of the classification.

A variety of shape feature descriptors have been proposed over the recent decades. In general, these descriptors can be categorized into three classes: 1) global descriptors; 2) local point descriptors; and 3) regional point descriptors. Since global descriptors, including extended Gaussian images (EGIs) [5], complex EGIs [6], and spherical attribute images [7], describe the overall shape statistic of an object, they are very useful in object retrieval and recognition. However, such global descriptors are sensitive to incomplete point clouds because they require the points of the objects to be well separated from the point cloud. Therefore, they are seldom employed directly in point cloud classification. On the other hand, the local point descriptors encode the local geometric properties of a point, including surface normal, surface curvature, and principal curvatures. The eigenvalue-based features are also taken as local point descriptors, which are widely used in classification [8], [9]. However, such local geometric properties are very sensitive

to noise. To address this problem, the regional point descriptors are presented to improve the robustness of a point descriptor, by aggregating the geometric properties of neighboring points. The spin image descriptor [10] combines the descriptive nature of the global object properties with robustness to partial views and clutters of the local shape descriptions, and it is widely used in classification and recognition [11]–[13]. In another approach, Frome *et al.* [14] presented two regional point descriptors: 1) the 3-D shape context and 2) the harmonic shape contexts. In that reference, it was shown that although in cluttered scenes the 3-D shape context descriptors are more stable than the spin image descriptor, they are much more computationally expensive. In general, it is difficult to select a proper support region and the raster size for these regional point descriptors, when the point density is varying.

Another methodology, which has been used in image classification, is the superpixel approach [15], [16], which aggregates the local neighboring points that have similar color and texture statistics. We apply a similar strategy to TLS point cloud classification. The TLS point clouds are adaptively clustered by the normalized cut into a few point clusters. The base unit for classification is a point cluster instead of a point. These adaptive clusters provide enough support for aggregation of the local geometric properties to form a robust feature descriptor. For example, Lim and Suter [17], [18] presented a supervoxel method for 3-D terrestrial range data classifications, in which they applied multiscale conditional random field methods to classify TLS point clouds. However, they only labeled some large objects, such as trees, grounds, and buildings, but did not try to label small objects (e.g., cars and people). Clearly, the rather small objects not only contain fewer points, which are often incomplete due to occlusions, but also keep on moving during the scan, and thus, the profiles of the point clouds are stretched. Obviously, it is very challenging to label such relatively small-sized objects. Patterson IV *et al.* [19] have recognized objects by using bottom-up and top-down descriptors. In [19], the EGIs [5] and spin image databases are used to implement the recognition. Following this method, the objects need to be manually labeled, and the recognition process is performed at the object level in their method. Golovinskiy *et al.* [20] first extracted small objects from the combination of mobile laser scanning and airborne laser scanning (ALS) point clouds. Then, they used a support vector machine classifier to classify the objects into semantic groups, such as cars, streetlights, trees, and fire hydrants. Although their approach can locate and segment small objects, their classification results are not very good because the shape features are not discriminative enough. Both the methods reported in [19] and [20] label objects at the object level. However, if the objects cannot be separated from the point clouds well, it is difficult to achieve good classification results with their methods. In contrast, we classify the point clouds on the point clusters with the help of our novel latent Dirichlet allocation (LDA)-based features, instead of relying on explicit object separation. Furthermore, as noted in [20], discriminative features are one of the keys to improve the classification results. Features used in [20], such as the number of points, that average the normal vector of the clusters may not be discriminative in some clutter scenes. For a discriminative cluster feature, it should not only generalize the features of points in a cluster but also contain the relation of the points. An approach is the so-called bag of words (BoW) [21], [22], which quantizes the point features into words and consolidates all point features of a point cluster into a word frequency vector, which is a sparse vector of occurrence counts of words. The BoW has been successfully employed in superpixel image classification, image retrieval, and text-document analysis. However, since the BoW ignores the relationship among words, the polysemies and synonymies of words cannot be captured. To fill this gap, based on the BoW, topic models have been developed to extract latent topics for the object classification and recognition. In the topic models, a topic level is added between the document and the word. A document is represented by a histogram of topics, while each topic represents a learned distribution of words. A famous topic model called probability latent semantic analysis (PLSA) [23] was presented to classify documents, and furthermore, the Dirichlet prior was added to the PLSA to generate the LDA [24]. The LDA was also introduced to classify images [25]. However, the flat LDA cannot discriminate the images well. The Chinese restaurant model [26] priors are added into the flat LDA to form hierarchical LDA (HLDA) in [27]–[29]. Although the HLDA can organize the latent topic into hierarchical tree structures to improve the precision of the image classification, its generation is usually slow and time consuming. Endres *et al.* [30] applied the LDA to unsupervised discover object classes in indoor scenes. Their method is designed for classifying objects that have been well segmented from the point clouds beforehand. However, segmenting the point clouds of TLS into independent objects is a difficult problem. In [30], the class number obtained by the unsupervised method is hard to correspond to the needed class number. Therefore, we only use the LDA to generate the feature of each cluster and apply the AdaBoost classifier [31] to classify point clusters. In this way, the class number can be corresponded to the required one.

Moreover, in order to account for the scale changes in the scanned point clouds, the point clouds are preprocessed into multiple scales during the training stage similar to the pyramid of scale-invariant feature transform (SIFT) [32]. Formally, we resample the point clouds into different scales and cluster the point clouds of each scale into different levels to form multiscale and hierarchical point clusters (MHPCs). The extracted cluster-based feature descriptors for each MHPC are scale invariant and robust against noise and incomplete point clouds. There are some other papers [33], [34] that focus on descriptions at different scales. In [33], Xiong *et al.* used point cloud statistics and relational information over fine (pointwise) and coarse (regionwise) scales. Xu *et al.* [34] employed three types of entities, such as single points, planar segments, and segments obtained by mean-shift segmentation, to classify the point clouds. In the aforementioned two methods, different scales are used for obtaining the context of the point cloud and obtaining the shape of the objects themselves. The MHPC feature can describe an arbitrary cluster itself in different scales, and it generalizes the features of points and the relations of the points in the cluster. Moreover, features such as number of points, standard deviation

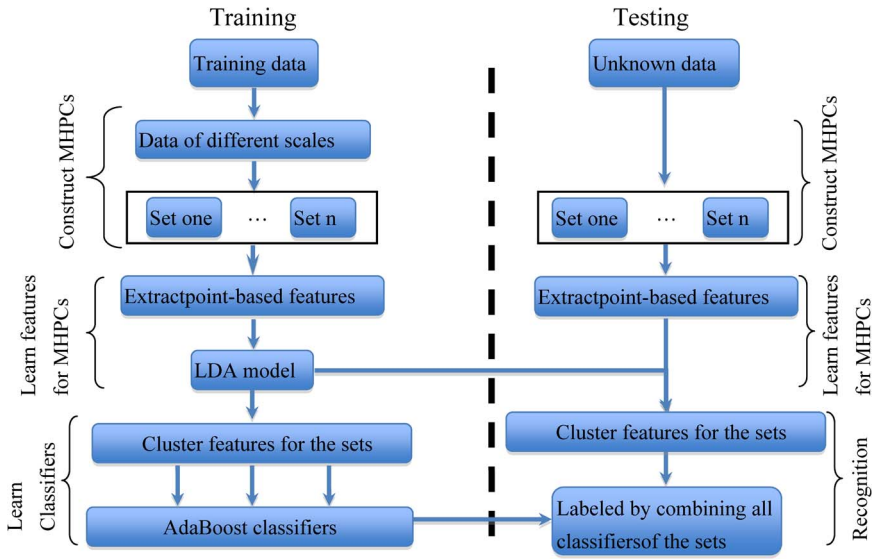


Fig. 1. Methodology for reconstructing 3-D tree models from the TLS point cloud.

in height for point clusters, and point-based features are still used in the aforementioned two methods, which are not robust and discriminative in cluttered scenes. Our MHPC features can be applied on the aforementioned two methods to improve the classification performance of points.

II. Overview and Contribution

Our goal is to extract robust and discriminative features for classifying noisy and cluttered urban scenes. Based on the extracted features, the point clouds can be robustly classified into semantic classes, such as buildings, trees, people, and cars. The contributions of our method are twofold.

1) A novel framework for effectively extracting the shape features of objects from complex TLS point clouds is presented. The MHPCs are proposed to be the basic units for extracting geometric features of the objects from TLS point clouds. We resample point clouds into different scales and form multiscale point clouds. The point clouds in each scale are aggregated into several point clusters with different sizes to form hierarchical point clusters. The point clouds of all scales in each level of the hierarchical point clusters are termed a point-cluster set. Compared with the point-based features, the robust features capturing the spatial relationships among the points in the point clouds can be derived from our point clusters. The point-cluster features are discriminative and robust against the local disturbances. Moreover, compared with the point cluster without scales and hierarchies, the features derived from the MHPCs are scale invariant and discriminative due to the explicit construction of hierarchical point clusters in the scale spaces. To classify a new point cloud, we train an AdaBoost classifier for each class on each point-cluster set using the prelabeled training data. Such classifiers on different levels for each class are combined to estimate the joint probabilities of assigning a class label to a point cluster over all levels. Our framework can effectively classify the objects in multiple scales with nonuniform point densities, particularly small objects, such as people and cars.
2) The LDA is applied to extract the features of the MHPCs in the point clusters. The LDA can extract discriminative and robust features with the same size from point clusters with different points; hence, the classifier can integrate the features of the MHPCs to directly classify point clouds conveniently. Applying the LDA reduces the sensitivity of the point-based features to imperfect point clouds with varying point densities, while maintaining the orientation invariance of the point-based features.

Our method has been tested on TLS point clouds acquired from three different urban scenes by single scans. Compared with other methods, the precision and recall generated by our method are almost the highest in classifying the four specific object classes, particularly for cars and people. Although semantic features are helpful to improve the classification accuracy, our aim is to obtain robust and discriminative shape features. Therefore, semantic features, such as the positions of objects and spatial relations among objects, are not employed in this paper.

The flowchart of our method is shown in Fig. 1. The remainder of the paper is structured as follows: The approach to construct MHPCs is described in Section II. The MHPCs containing multiple scales and levels are clustered. Finally, the MHPCs are combined into some point-cluster sets whose count is the same as the count of the levels. In Section III, the LDA-based cluster feature is described. The point-based features are firstly obtained in Section III, including the eigenvalue-based features and spin image. The features are the base for deriving the feature of an MHPC. In Section III, the LDA model is trained by using the point-based features of the MHPCs. The obtained LDA model is then employed to extract the features of the clusters. In Section IV, the whole supervised classification process is described. The cluster features in each point-cluster set are used to train the AdaBoost classifiers. The probabilities of the point-cluster sets obtained by AdaBoost classifiers are combined to implement the classification. In Section V, the

experimental results of the TLS point cloud classification are performed to validate our method. We also compare our method with the supervised LDA (sLDA) method [35] and methods using other features, such as the point-based features, the features obtained by the BoW, and PLSA.

III. Construction of MHPCS

The point-based features only describe the shape characteristic of a small region. For example, the points on a tree crown are cluttered, and the points on a wall are almost distributed on a plane. The point-based features can be used to distinguish them. However, for another example, a small part of the tree stems and car exteriors are relatively piecewise flat, and at these regions, the point-based features can hardly distinguish them. To classify the point clouds of the regions where the points are difficult to be distinguished, we aggregate the points into point clusters that provide enough support for extracting robust features, while being small enough, without containing multiple objects. We take such a point cluster as a basic unit for the object classification. Moreover, to extract more discriminative shape features from the point clusters, the point clusters are resampled into several scales and aggregated into several levels. Formally, we name such point clusters as MHPCs. The MHPCs are constructed by the following five steps.

1) Following the previous method [20], the isolated points and ground points are first removed. The removal of the ground points helps to determine the connectivity of objects.
2) To make the obtained shape features insensitive to point density, the features should be scale invariant. Similar to the image pyramid, point clouds are resampled into several scales. More specifically, for a point cloud at the ith scale, we resample it to make the point density in the $(i+1)$th scale be a percentage of the point cloud in the ith scale. The resampling procedure does not stop until the point density is less than 50% of the angular resolution of the scanner. According to the Shannon sampling theorem, if the point density is too low, the shape features obtained from the sparse resampling points cannot preserve the raw surface structure. On each scale, the following clustering steps (steps 3–5) are processed simultaneously.
3) In the nonground point clouds, we search the k_1 (k_1 is an integer) closest points of each point and then connect the point with its k_1 closest points by edges. In this way, an undirected graph $G_1(\mathbf{V}, \mathbf{E})$ is generated, where $\mathbf{V}$ is the set of the points and $\mathbf{E}$ is the set of the corresponding edges. The Euclidean distance between two connected points is taken as the weight of the edge. After G_1 is generated, all of the connected components of G_1 can be found.
4) Since objects are often close to others in cluttered urban scenes, a connected component can contain more than one object. In a connected component, a local maximum point may represent the top of an object. To further break the connected component into smaller pieces so that single objects can be isolated, a moving window algorithm is applied to search the local maximum points in a 2-D raster image, which 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 maximum points are found, the graph cut [36, Eq. 1] is employed to segment the connected component, and the local maximum points are taken as seeds. After the graph cut is performed, the connected component is divided into several point clusters

$$E(\mathbf{L}) = \sum_{i \in \mathbf{V}} D_i(l_i) + \sum_{(i,j) \in \mathbf{E}} V_{ij}(l_i, l_j) \tag{1}$$

$$D_i(l_i) = \begin{cases} 1, & \text{if } p \text{ is not a seed} \\ 0, & \text{if } p \text{ is a seed} \end{cases} \tag{2}$$

$$V_{ij}(i,j) = \exp(-d_{ij}) \tag{3}$$

where l_i is the label of point i, the number of the labels equals to the number of the seeds, $E(\mathbf{L})$ is the total energy of all the points with the corresponding labels, $D_i(l_i)$ is the data cost of point i, $V_{ij}(i,j)$ is the smooth cost of points i and j, and d_{ij} is the Euclidean distance between points i and j.
5) Each of the point clusters acquired by the aforementioned steps 3–4 may still contain more than one object. To achieve discriminative cluster features, a cluster should contain only one single object (or a part of it). Furthermore, the distribution of points on the object surface should be as even as possible. Motivated by the fact that the normalized cut [37] can aggregate the points with uniform distribution into one cluster, it has been employed here to partition a large point cluster into two new clusters under the condition that the number of points in the cluster is larger than a predefined threshold δ_m. To ensure that a point cluster contains enough feature information, δ_m is determined by the angular resolution. The higher (lower) the angular resolution is, the larger (smaller) the value δ_m takes. However, labeling a point cluster with only one size may still be difficult. To address the problem, the point clouds are split hierarchically. The point clusters with the smallest size δ_m are the lowest level, i.e., the nth level. The size of the jth $(j < n)$ level above it is set to $(n - j + 1) \times \delta_m$. Fig. 2 shows an example of the hierarchical point clusters with three levels. The hierarchical clusters mean that the features can better recognize the objects from the scanned scenes with different sizes. For example, Fig. 3(b)–(d) shows the three levels of point clusters of a car illustrated by the red points in (a). At the first level, there are trihedral angles in the point clusters colored in red. The cluster may be labeled as cars. At the second level, there are only some shape features containing folds or bulges in the regions of the point clusters colored in red, but the red region is somewhat near flat. The cluster may be labeled as cars or buildings. At the third level, the red region is almost flat, and the cluster may be labeled as cars or buildings. Combining the shape features of the three levels can help the cluster colored in red in the third level to be labeled

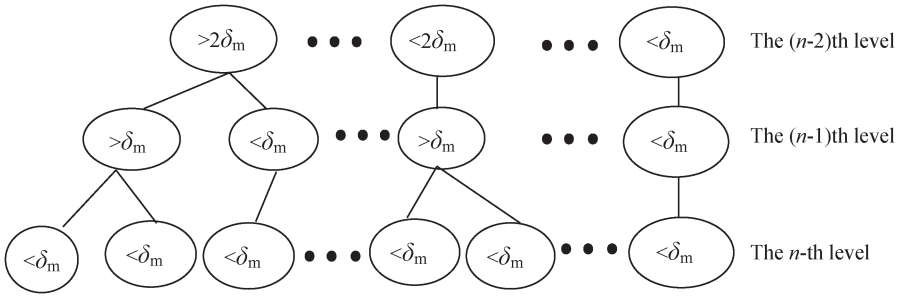


Fig. 2. Illustration of a three-level hierarchical point cluster.

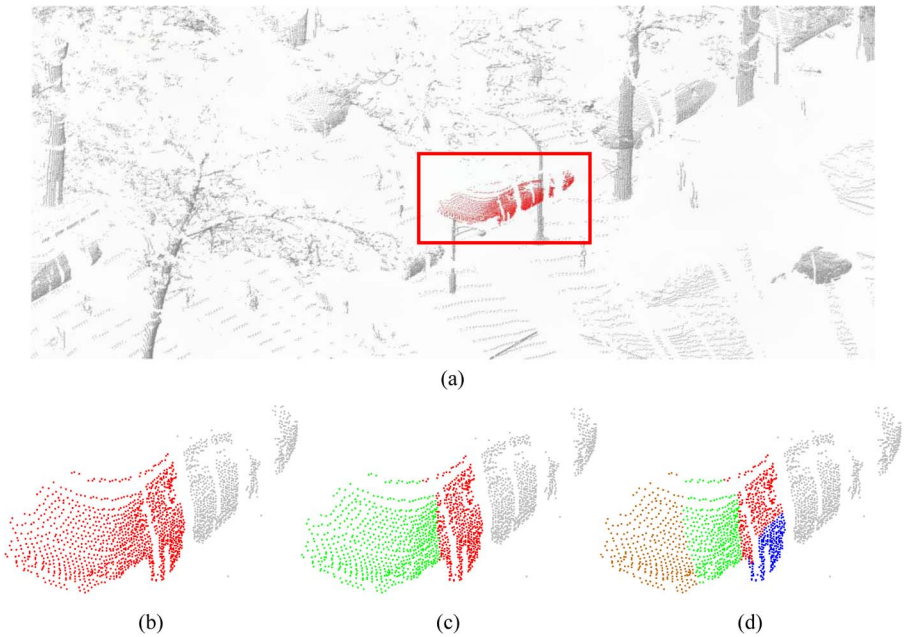


Fig. 3. Hierarchical point clusters of a car. (a) Point clouds of a scene in which a car is colored in red. (b) The point clusters of the car in the first level. (c) The point clusters of the car in the second level. (d) The point clusters of the car in the third level. Different colors indicate different clusters.

correctly because the cluster in the first and second levels can be labeled precisely.

Finally, we term the point clusters at each level of all scales as a point-cluster set; hence, there are totally n point-cluster sets. The set i contains the point clusters in the ith level of all scales. The MHPCs are constructed in this way to ensure that the extracted features are scale invariant and discriminative in the later procedures.

IV. Extraction Features for MHPCS

Here, we further describe the feature extraction for MHPCs consisting of two steps. The point-based features are first extracted from the raw point clouds. Next, the LDA model is trained using the point-based features of the MHPCs, and the feature of each MHPC can be obtained by the LDA model.

A. Point-Based Features

Although the point-based features are sensitive to noisy and cluttered point clouds, the LDA model can extract insensitive features of the MHPCs from the point-based features in a point cluster. The point-based features are the base for constructing the features of the MHPCs. Since the point clouds have been resampled into multiple scales, the point-based features are correspondingly extracted in each scale.

Let $\mathbf{N}_p = \{\mathbf{q} | \mathbf{q}$ is one of the k_2 closest points of $\mathbf{p}\}$ be the set of points, which are the k_2 nearest neighborhoods of p, and denote it as the support region of $\mathbf{p}$. Let us also assume $\overline{\mathbf{p}}$ to be the centroid of all points in $\mathbf{N}_p$. In order to guarantee that the point density in the support region is uniform, the value of k_2 cannot be too large, and it depends on the angle resolution of the point clouds. Otherwise, the points in $\mathbf{N}_p$ may be distributed on different parts of an object or even on different objects. For example, assuming that $\mathbf{p}$ is located on a building far from the scanner during the scan, some of the points in $\mathbf{N}_p$ may be distributed on building walls and others on balconies, which makes the shape features of the points unstable.

After the support region is determined, the features obtained by the eigenvalues and spin image descriptor [10] are extracted.

On the one hand, the features of $\mathbf{p}$ obtained by eigenvalues represent the distribution of the points in $\mathbf{N}_p$. The eigenvalues λ_1, λ_2, and λ_3 $(\lambda_1 \geq \lambda_2 \geq \lambda_3)$ are obtained by the following covariance matrix $(\mathbf{C}_p)$ of $\mathbf{N}_p$:

$$\mathbf{C}_p = \frac{1}{|\mathbf{N}_p|} \sum_{\mathbf{q} \in \mathbf{N}_p} (\mathbf{q} - \overline{\mathbf{p}})(\mathbf{q} - \overline{\mathbf{p}})^T \tag{4}$$

where T is the transpose operator.

The ranges of the eigenvalues computed for covariance matrices of different points are different. To compare these eigenvalues of different covariance matrices, the eigenvalues of each covariance matrix need to be normalized. That is

$$\lambda_i = \frac{\lambda_i}{\sum_i \lambda_i} \quad i = 1, 2, 3. \tag{5}$$

The features obtained by eigenvalues are combined into a 6-D vector as follows [38]:

$$\mathbf{F}_{\text{eigen}} = \left[\sqrt[3]{\prod_{i=1}^{3} \lambda_i}, \frac{\lambda_1 - \lambda_3}{\lambda_1}, \frac{\lambda_2 - \lambda_3}{\lambda_1}, \frac{\lambda_3}{\lambda_1} - \sum_{i=1}^{3} \lambda_i \log(\lambda_i), \frac{\lambda_1 - \lambda_2}{\lambda_1} \right]$$

which are the structure tensor omnivariance, structure tensor anisotropy, structure tensor planarity, structure tensor sphericity, structure tensor "eigenentropy," and structure tensor linearity, respectively.

On the other hand, a spin image can capture the majority of local shape information presented in a 3-D scene. It is a 2-D parameter space histogram. Following the definition of the spin image, each 3-D point can be projected onto a 2-D space with the x- and y-axes. We define the spin image as $\mathbf{F}_{\text{spin}}$, taking the normal vector of a point as the rotation axis and setting the size to 3×4 bins. Therefore, $\mathbf{F}_{\text{spin}}$ has twelve values. Since there are a few points in $\mathbf{N}_p$ to avoid excessive zero values in the spin image, we compute the absolute value of the negative values on the y-axis to obtain their corresponding positive values. In this way, the range of the y-axis is from zero to the infinite, and the values of the rasters on the positive y-axis increase. In the spin image of p, the raster size along the x-axis equals to a certain value (e.g., one third of the Euclidean distance between the farthest point in $\mathbf{N}_p$ and p). Along the y-axis, there are also multiple scales. In this paper, the y-axis is divided into four scales: the first scale is 0–0.02 m; the second scale is 0.02–0.04 m; the third scale is 0.04–0.06 m; and the last scale is 0.06 m to infinity. The points in $\mathbf{N}_p$ fall in the rasters, and the number of the points falling in each raster is accounted. The twelve values of $\mathbf{F}_{\text{spin}}$ in the spin image and the six values of $\mathbf{F}_{\text{eigen}}$ are combined into an 18-D descriptor for p as $[\mathbf{F}_{\text{spin}}, \mathbf{F}_{\text{eigen}}]$ defined as $\mathbf{F}_{\text{point}}$. $\mathbf{F}_{\text{point}}$ inherits the orientation invariance of $\mathbf{F}_{\text{spin}}$ and $\mathbf{F}_{\text{eigen}}$.

B. Extracting Features of MHPCs

After $\mathbf{F}_{\text{point}}$ of all points and MHPCs are obtained, one feature vector is extracted from $\mathbf{F}_{\text{point}}$ of points of each MHPC. The feature cannot easily average $\mathbf{F}_{\text{point}}$ of the points but should generalize them. Although the number of the points in the different clusters varies, the length of the feature of each MHPC should be the same, so that the features can be conveniently employed by the classifiers. The LDA extracts vectors with the same length from various clusters. Each vector generalizes $\mathbf{F}_{\text{point}}$ of points and contains the relation of points in each cluster. Moreover, the features should inherit the advantage of $\mathbf{F}_{\text{point}}$ and reduce its sensitivity to the varying point density.

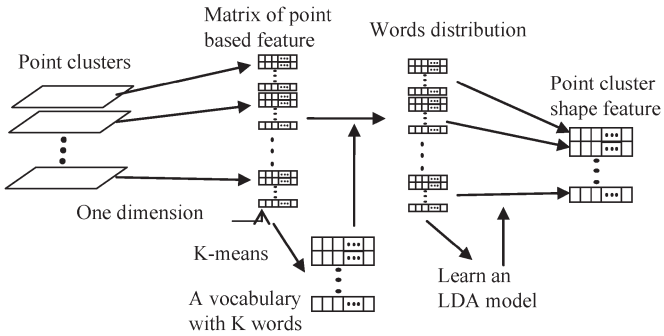


Fig. 4. Process of extracting the features of the MHPCs.

The LDA was originally proposed to classify documents of different length and words. It extracts a certain number of latent topics to describe the documents. Each document is expressed by a vector that consists of the probability of each latent topic. In this paper, we apply the LDA to extract the point clusters' features. A vector expressed by the LDA can be a feature of a point cluster. More concretely, we define the document, document set, dictionary, and word. Each MHPC is taken as a document, and MHPCs as a set of documents. The dictionary and words are obtained by vector quantization. The K-means is employed to aggregate all $\mathbf{F}_{\text{point}}$ into K clusters and obtain the center vectors of the clusters. The K center vectors are the words, and the K words form the dictionary. $\mathbf{F}_{\text{point}}$ of a point can be represented by a word, if the distance between it and the word is the shortest in all of the words of the dictionary. All $\mathbf{F}_{\text{point}}$ in an MHPC are replaced by the words in the dictionary. Then, the appearance frequency of each word in an MHPC is computed to generate a word frequency vector with the length of K. The word frequency vectors of all MHPCs are derived. The LDA model is trained by using them.

The probability vector of latent topics in each MHPC is obtained by the LDA model. It is the feature of the MHPC. Fig. 4 demonstrates the process for extracting this feature. The distance between the word and $\mathbf{F}_{\text{point}}$ is computed during the process. Since the data ranges of the different dimensions are not the same, the distance is determined by the dimensions with large data ranges. Therefore, each dimension of all $\mathbf{F}_{\text{point}}$ vectors needs to be normalized, and then, the features can be used for vector quantization. Otherwise, the result of vector quantization is not satisfactory, and the features of the MHPCs are also not good.

After the LDA model has been learned, the feature of the MHPC is extracted by the following steps. $\mathbf{F}_{\text{point}}$ of points in this cluster are computed, and then, they are normalized in the same way as during the LDA training process. These normalized $\mathbf{F}_{\text{point}}$ are represented by words of the dictionary. The word frequency vector of the cluster is calculated. The LDA model works on the word frequency vector to obtain a feature of the MHPC.

In the process for extracting the features of the MHPCs, the LDA model does not change the orientation invariance of the point-based features, and it is trained from the MHPCs as all of the scales and levels of the point clusters are employed, keeping

TABLE I
SUMMARY OF THE THREE DATA SETS AND THE GROUND TRUTH LABELS

	Scene I	Scene II	Scene III
Scanners	RIEGL LMS-Z620	RIEGL vz-1000	RIEGL vz-1000
Measurement principle		Time of flight	
Measurement Range(m)	2-2000	2.5-1400	
Accuracy	10mm	8mm	
Precision	5mm	5mm	
Beam Divergence	0.15mrad	0.3mrad	
Laser wavelength	Near infrared	1,550nm	
Horizontal and vertical angle spacing	0.57°	0.60°	0.60°
People points	109,233	63,496	32,814
Tree points	1,360,821	494,430	747,541
Building points	250,094	501,389	188,172
Car points	34,440	197,261	74,282
Total points	1,754,588	1,256,576	1,042,809

the scale invariance. Therefore, our obtained features contain two invariants. The feature of the MHPC is the probability vector of each latent topic; hence, the main latent topics that can represent the main features of a cluster have a high probability. Noise and the features that do not describe the object surfaces are well presented by the latent topics with a small probability. The features are able to extract the common parts of the cluster and reduce the sensitivity of the point-based features. In this way, the dimension of the features of the MHPCs can be determined by the users.

V. MHPC CLASSIFICATION WITH LDA-BASED FEATURES

We use the AdaBoost classifier to classify the MHPCs into four classes (people, trees, buildings, and cars) by using the LDA-based features. AdaBoost is a supervised learning algorithm that combines weak classifiers into a much stronger classifier. It is more resistant to the overfitting than many other machine learning algorithms, as pointed out in [2] and [39], and performs well, as shown in [40]. Since our classification is supervised, the training data are clustered to the MHPCs, and then, the LDA-based features are extracted. To prevent the training results from being affected by the clutter points, the clusters that contain less than 20 points are removed. After all of the features of the MHPCs are derived by the LDA model, the one-versus-all AdaBoost classifiers are trained. For n point-cluster sets, there are correspondingly $4n$ AdaBoost classifiers, as four classifiers are trained for the four classes of each set. The LDA model and AdaBoost classifiers are obtained in the training process. Then, they are applied to classify the unlabeled point clouds.

Given the test point clouds, the data set is also clustered to the MHPCs but without being resampled. That is, there are still n levels with one scale. The LDA-based features of the MHPCs are extracted. Then, the AdaBoost classifiers classify the point clusters in each set. The clusters that contain less than 20 points are not classified and are grouped into the classes that the nearest clusters belong to.

Following [16], the probability of assigning a label l_i to a specific cluster can mathematically be expressed as

$$P_{\mathrm{num}}(l_i, \mathbf{F}) = \frac{\exp\left(H_{\mathrm{num}}(l_i, \mathbf{F})\right)}{\sum_i \exp\left(H_{\mathrm{num}}(l_i, \mathbf{F})\right)} \tag{6}$$

where $\mathbf{F}$ is the features of the MHPCs; num is an integer $(1 \leq num \leq \mathrm{n})$; $H_{\mathrm{num}}(l_i, \mathbf{F})$ is the output of the AdaBoost classifier for l_i.

The point clusters are then finally labeled by the joint probability of assigning a label over the multiple point-cluster sets. During the training process, the training data are manually labeled carefully; hence, each cluster only contains one specific object class. In the generalization process, a cluster in the point-cluster sets at higher levels may contain more than one specific object class; therefore, we only label the clusters in the lowest level point-cluster set. To achieve the scale-invariant classification, we combine the shape features of all point-cluster sets together. The joint probability for labeling l_i to a cluster in a point-cluster set can be mathematically obtained as

$$P(l_i) = \prod_{\mathrm{num}=1}^{n} P_{\mathrm{num}}(l_i, \mathbf{F}). \tag{7}$$

Finally, all point clusters in the lowest level point-cluster set are labeled by the highest probability of the labels.

VI. RESULTS AND DISCUSSION

To validate the performance of our method, we perform both qualitative and quantitative evaluations on three scanned urban scenes. The point clouds of the three scenes were obtained by a TLS in a single scan. The majority of objects that appear in the urban scenes are buildings, trees, cars, and people. Point density of the point clouds in the three scenes varies a lot, with respect to the distance of the scanned objects to the scanner. Moreover, the point clouds of the objects acquired by the single scan are often incomplete due to occlusions. For example, the point clouds of the faraway objects are sparse and fragmented because the laser is blocked by the closer objects. To train our classifiers and evaluate our methods quantitatively, we further assigned the ground truth labels on all three data sets manually. The statistics of our data and the ground truth labels are shown in Table I. In the experiments, we manually set the number of the scales and levels of the point clusters to three, and the percentage is set to 75% empirically. If sampling with more scales and levels exceeds the angular resolution of the used scanner, then it does not guarantee enough points being sampled for small objects.

We implemented the proposed method in C++. The algorithm runs on a computer with an Intel Core i7-4770 K processor at 3.50 GHz and 8-GB RAM. It takes about 40.8 min to learn the LDA models and AdaBoost classifiers. It takes approximately 52.4, 20.6, and 23.6 min to classify the point clouds in Scenes I, II, and III, respectively. During the processing, feature extraction and word quantization cost about 64% of the whole running time. However, this process and most of these steps are parallelizable. Therefore, they can be implemented by employing a parallelization scheme to further reduce computing time.

The quantitative evaluations consist of two parts: 1) comparison of our method with other methods by precision/recall; 2) the sensitivity to varying point densities and different parameters, including the number of words, latent topics, levels, and scales.

A. Comparison With Other Methods

To show the advantage of our method, we compare our method with three methods that use different features and with the sLDA method on point cloud classification in both learning and generalization processes. The first method (Method I) uses the point-based features introduced in Section II. It directly applies the point-based features to classify point clouds without aggregating the data set into clusters. Hence, Method I does not have hierarchical structures. The second method (Method II) employs the feature constructed by the BoW instead of our LDA-based features for the MHPCs. The third method (Method III) uses the feature obtained by PLSA instead of the LDA. The number of words and the topics used in PLSA and sLDA are the same as in the LDA. For the sLDA method, we only train one sLDA model and use it to classify the clusters in the third point-cluster set. The aforementioned two methods only have one learning result for each object class. For the quantitative evaluation, we train the models of our method and four other methods from the training data, respectively. Here, the models that obtain the best results by the combinations of scales and levels in the generalization process are applied in all the five methods. The sLDA method is trained on the clusters of the first and second scales and three levels. The other methods are executed on all of the clusters, and each method achieves a learning result for each point-cluster set.

We selected some of the points from the scenes as the training data set, which is illustrated in Fig. 5. The whole training data set contains 647 635 points, including 67 604 people points, 303 994 tree points, 141 820 building points, and 134 217 car points. The points of trees and buildings in the training data set are mainly obtained in Scene I. The points of people in the training data set are mainly obtained in Scene II. The points of cars in the training data set are mainly obtained in Scene III. Some particular objects are also selected, e.g., the ladders classified as building only appear in Scene II, and the baby carriages classified as cars only appear in Scene I. To validate the performance of the methods in terms of precision/recall, Table II shows detailed learning performances of the five methods. As shown in Table II, precision and recall of all of the objects achieved by using our method are higher than the other four methods. The LDA-based feature describes the training data best.

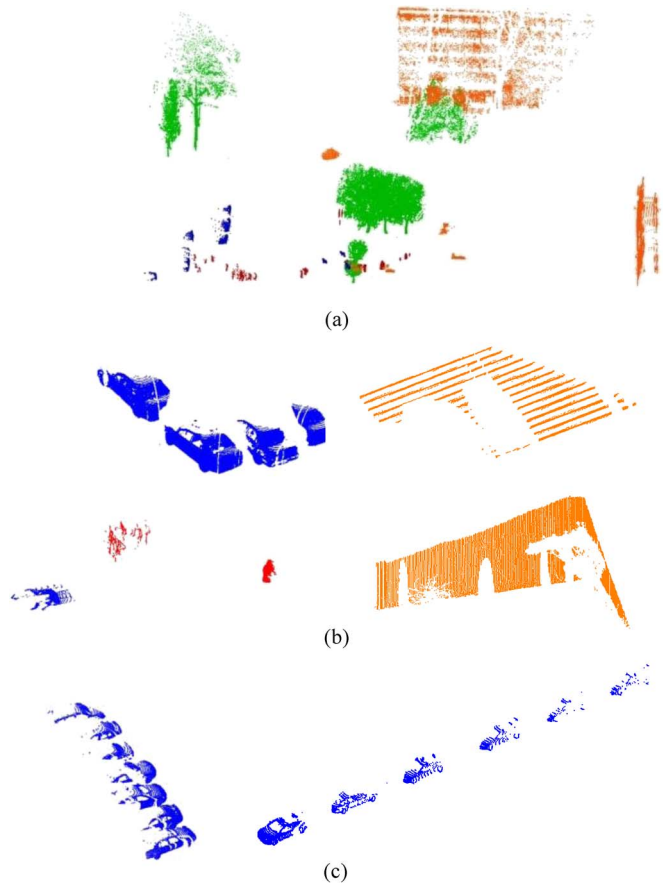


Fig. 5. Training data. Green points represent trees, orange points are buildings, blue points are on cars, and red points are on people. (a) Part of the training data set obtained from Scene I. (b) Part of the training data set obtained from Scene II. (c) Part of the training data set obtained from Scene III.

The generalization abilities of the above methods have also been tested. Table III demonstrates the performance of the methods for the three scenes in terms of precision/recall and accuracy. Figs. 6–8 show the classification results of the methods visually.

As shown in Table III, the precision and recall of most classes by our method are fine. This means that we can distinguish these types of objects very well. Although some parts of these objects, such as tree stems, building corners, and cars, are difficult to be distinguished, most of them are correctly recognized by our method. The classification precision of people is lower than that of other types of objects. People are usually in motion in the scanned scenes, and their size is relatively small. Therefore, only very few points are acquired for them, and the features of people are hard to be obtained. Moreover, the features for people with different postures may not be learned from the training samples; hence, they are difficult to be recognized. The number of points on cars is very few. If a very small part of the other classes with many points is recognized incorrectly, the precision is low. The recall of cars in Scene I is low. This is because, in Scene I, there are six baby carriages, which are also grouped into the car class. However, they are

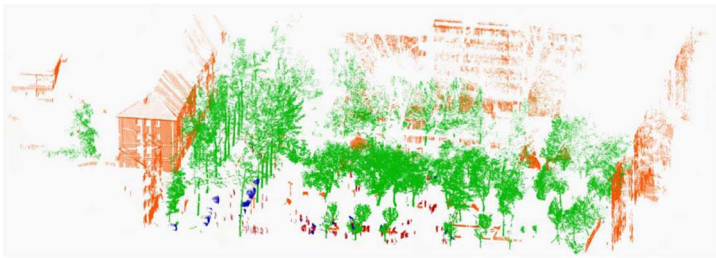
TABLE II
PRECISION/RECALL OF DIFFERENT METHODS IN THE LEARNING STAGE

	Point-Cluster sets	People (%)	Trees (%)	Buildings (%)	Cars (%)
Our method	1	98.9/95.7	98.2/99.5	97.6/97.9	98.8/96.8
	2	1/97.3	99.1/99.7	98.4/99.3	99.4/98.0
	3	1/97.4	99.9/1	99.5/1	99.2/99.6
Method I		62.6/54.1	81.8/88.2	75.1/77.1	65.0/62.0
Method II	1	89.4/52.1	82.7/99.4	80.2/69.6	89.7/66.2
	2	75.0/27.8	79.6/99.2	77.5/76.5	90.1/61.4
	3	90.1/51.4	83.0/99.5	83.2/67.9	85.5/64.0
Method III	1	71.6/41.9	73.9/92.6	69.8/52.9	78.7/62.9
	2	71.5/49.9	70.5/97.1	73.9/32.3	89.0/54.4
	3	82.0/37.0	77.5/93.5	66.8/51.6	76.4/63.1
sLDA		78.8/66.9	91.2/96.6	82.8/80.9	78.7/73.1

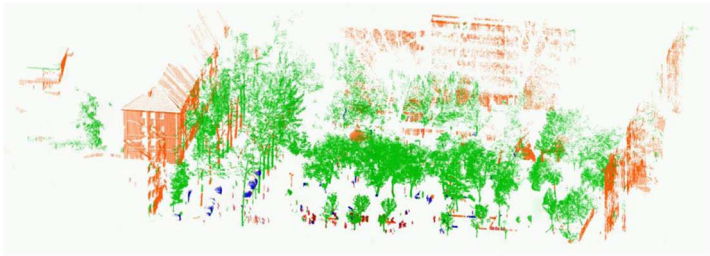
TABLE III
PRECISION/RECALL AND ACCURACY OF DIFFERENT METHODS ON THE THREE SCENES IN THE GENERALIZATION STAGE

Scene I	People (%)	Trees (%)	Buildings (%)	Cars (%)	Accuracy (%)
Our method	82.9/62.7	95.4/98.3	89.9/86.7	52.9/45.4	93.4
Method I	28.6/32.5	89.5/87.4	61.3/62.0	9.1/12.8	78.9
Method II	81.6/52.6	94.6/98.0	87.8/88.2	50.2/33.4	92.5
Method III	32.2/12.5	84.0/95.2	60.0/41.0	0/0	80.4
sLDA	68.4/33.5	91.9/97.7	84.1/80.7	42.9/18.1	89.7
Scene II	People (%)	Trees (%)	Buildings (%)	Cars (%)	Accuracy (%)
Our method	78.8/77.5	95.9/90.1	89.0/93.3	83.7/86.4	90.2
Method I	53.8/54.8	79.6/86.2	84.4/79.0	63.0/59.8	77.6
Method II	70.9/81.1	93.0/90.8	92.8/89.2	80.3/89.6	89.4
Method III	68.8/56.2	89.0/91.2	82.9/91.2	82.0/65.8	84.6
sLDA	66.8/50.5	94.6/90.2	84.8/94.1	79.7/72.9	87.0
Scene III	People (%)	Trees (%)	Buildings (%)	Cars (%)	Accuracy (%)
Our method	84.9/69.4	98.2/95.6	83.7/92.3	77.4/85.6	93.3
Method I	56.9/47.4	92.2/85.1	56.5/73.2	50.5/55.1	79.7
Method II	56.8/75.2	98.2/95.0	88.4/90.4	78.1/88.0	93.0
Method III	71.7/33.7	91.2/94.9	75.7/79.1	72.3/51.4	87.0
sLDA	84.3/63.6	95.2/96.4	81.6/87.7	74.9/58.4	91.3

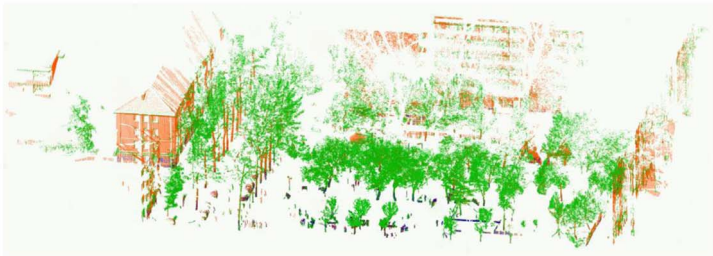
not similar to cars in shape. There are 9762 real car points and 24 678 baby carriage points in the car class of Scene I. Although two baby carriages were put into the training data, they are only a small part of the training data, and it would not be very helpful for recognizing the baby carriages. However, if the points on the baby carriages are removed, then most of the real car points are recognized, as shown in Scene I, by using our method in Fig. 6. As listed in all scenes, by using our method in Figs. 6–8, the visual confirmation indicates that our method achieves similar recognition rates to the one recorded in Table III, for all of



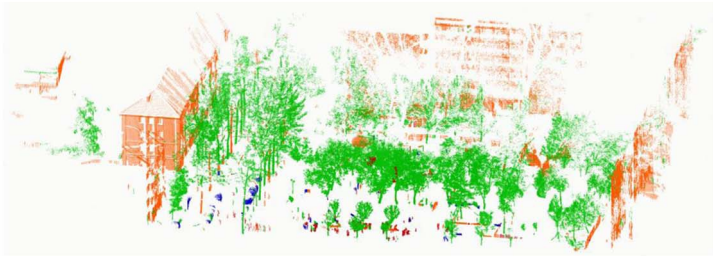
Ground truth.



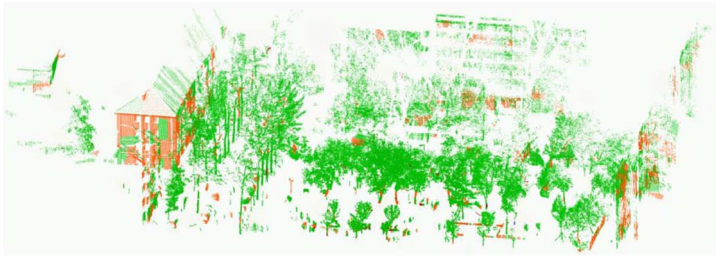
Out method



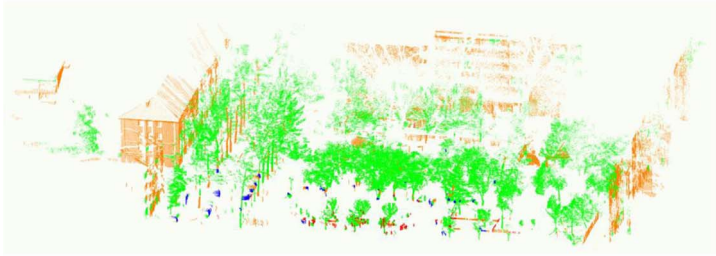
Method I



Method II.



Method III.



sLDA.

Fig. 6. Results obtained by different methods in Scene I. The points on trees, buildings, cars, and people are colored in green, orange, blue, and red, respectively.

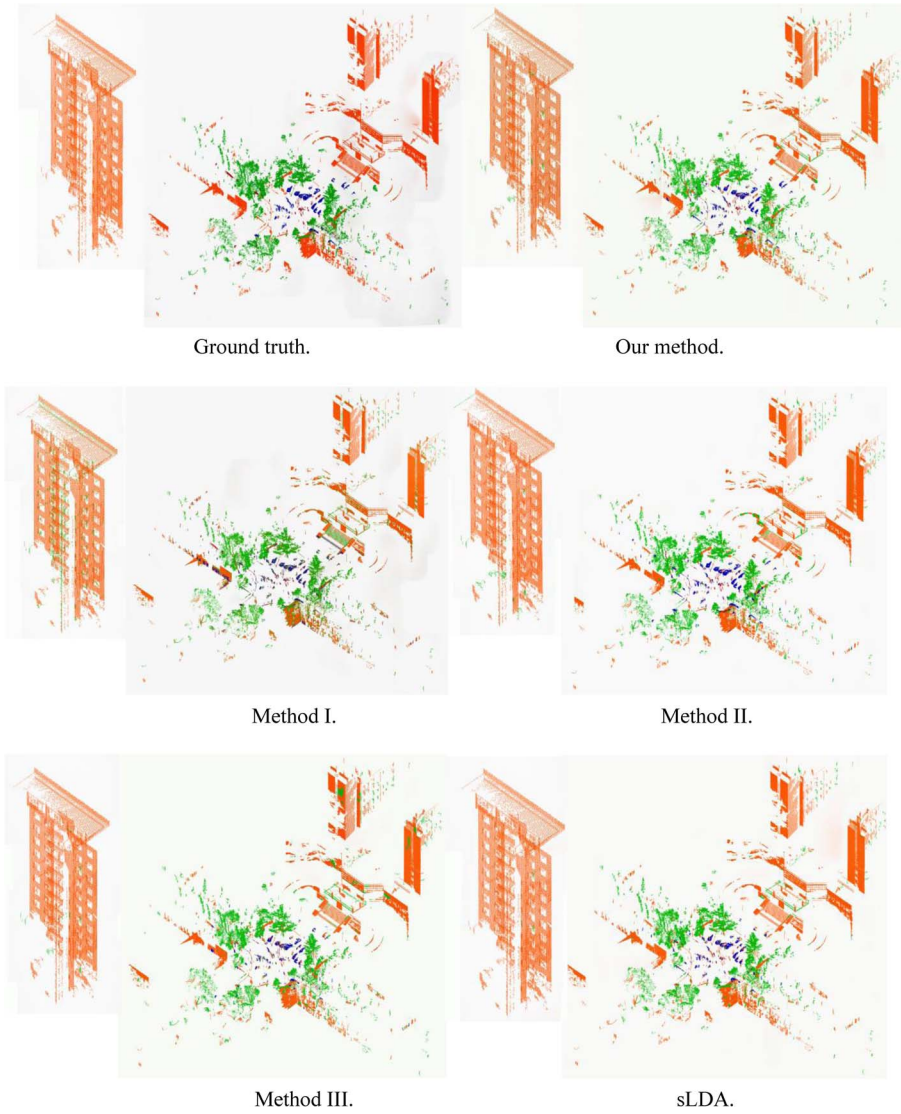
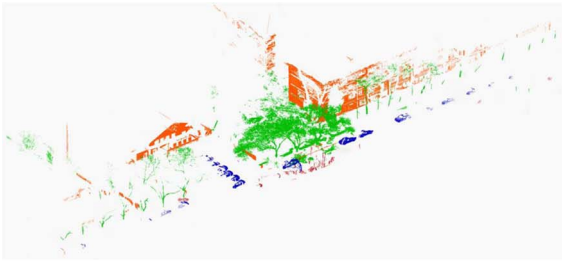


Fig. 7. Results obtained by different methods in Scene II. The points on trees, buildings, cars, and people are colored in green, orange, blue, and red, respectively.

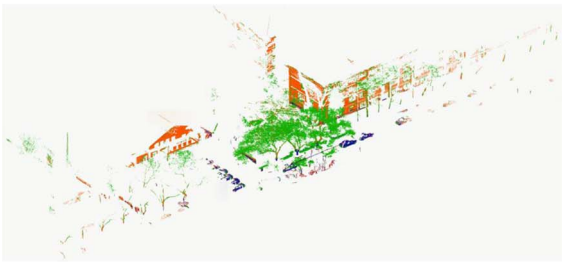
the classes of the three urban scenes. Compared with other methods, we note that the accuracy of our method is the highest. The result of Method I in all of the scenes in Figs. 6–8 is the worst, which means that the use of the cluster as the basic unit is helpful for improving the classification results. Due to the fact that it is point based, the result looks cluttered. The different classes are mixed together even in a small region. The classification of our method is better than Method III, which means that the Dirichlet prior performs well. The result is illustrated in each scene in Figs. 6–8. The buildings are labeled as trees in some places obviously, which corresponds to the low recall of buildings. The sLDA method achieves a good classification performance for trees and buildings, but it cannot effectively recognize people and cars from the cluttered point clouds. This may be because the sLDA method only classifies the deepest level of the point-cluster set. The shape features of people and cars are similar to other object classes in a small area, and they cannot be recognized well if the shape features of the other two point-cluster sets are not used. It is hard to classify the clusters by combining three sLDA models. Our method can easily incorporate the shape features of the three point-cluster sets by (7) to achieve high classification accuracy. This indicates that the hierarchical clusters are helpful for improving the classification accuracy of the small objects. As shown in the result obtained by the sLDA method in Figs. 6–8, many car points are labeled as other classes by mistake. Method II performs well in classifying trees, buildings and cars. The number of points on people varies in different clusters; hence, the word frequency vectors contain a different number of words. This makes the feature not well suitable for classifying people. In Figs. 6–8, although the results of trees, buildings, and cars recognized by Method II are similar to those recognized by our method, people are not recognized well. That indicates that our method is more suitable for recognizing the small objects. The precision/recall diagram and Figs. 6–8 illustrate that our method is highly accurate with good performance values.

B. Sensitivity of Our Method

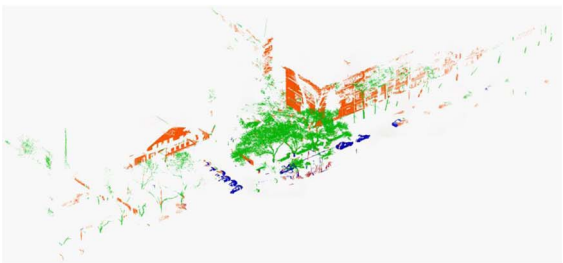
Here, we analyze the impact of the parameters, including the topic number and word number, and the number of levels and scales, on the classification.



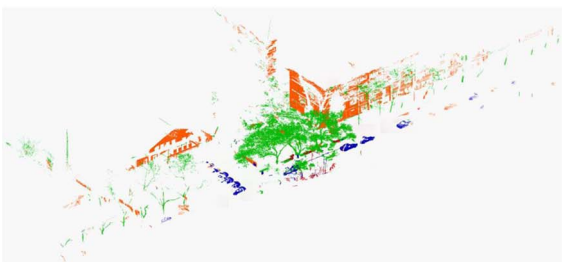
Ground truth.



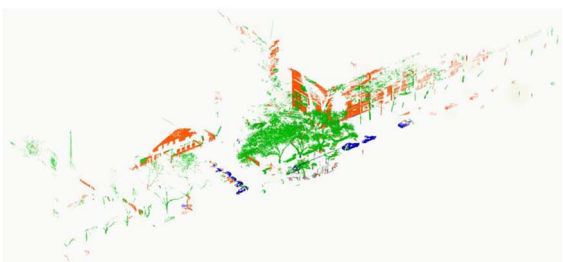
Out method.



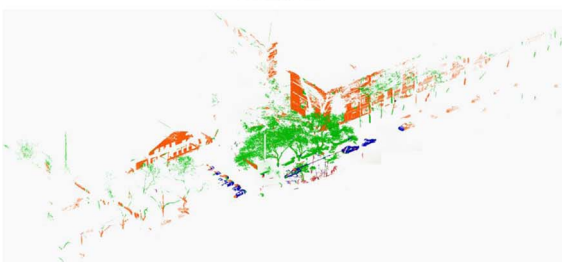
Method I.



Method II.



Method III.



sLDA.

Fig. 8. Results obtained by different methods in Scene III. Points on trees, buildings, cars, and people are colored in green, orange, blue, and red, respectively.

We classify the point clouds with different parameters. Fig. 9 shows the classification quality of Scenes I, II, and III represented by using the F_1 measure (8) respectively corresponding to the number of topics and words. That is

$$F_1 = \frac{2(\text{recall} \times \text{precision})}{\text{recall} + \text{precision}}. \tag{8}$$

The aforementioned performance evaluation is the result of using 300 words and 30 topics. We respectively select 100, 200, 300, 400, and 500 words and 30 topics to analyze the impact of the word number on classification quality and 20, 25, 30, 35, and 40 topics and 300 words to show the impact of the topic number on the classification quality. In Fig. 9(a)–(c), it can be observed that the word number and topic number have slight influence on the quality of the four classes in the three scenes. Most of the differences of F_1 of the classes are all below 5%, which means that our framework can achieve very good performance with very few words and topics. Although the differences of F_1 of the cars in Scene I and the people are also small, F_1 changes when the parameters vary. The graphs of the other three classes in Fig. 9 are almost straight lines. Features on the surfaces of trees, buildings, and cars may be simple and obvious. The selection of only a few words and topics is enough to describe these different features. For cars in Scene I, since there are not enough baby carriages samples here, the words cannot describe the baby carriages well. People are often in motion, and the scanned points on the distant people are few. Therefore, the word number and topic number affect the recognition of people more than other objects.

To evaluate the effectiveness of the MHPCs and the robustness of our method against varying point density, the results of the three scenes with 50%, 60%, 70%, 80%, 90%, and 100% of the raw points using different level and scale combinations are illustrated in Figs. 10–12, respectively. For simplicity, we term the three scales and three levels, three scales and one level, one scale and three levels, and one scale and one level as $3s3l$, $3s1l$, $1s3l$, and $1s1l$, respectively. The data set with one scale is the raw point cloud. The point clusters with one level are generated by dividing the point clouds until point clusters are less than 500 points, i.e., the minimum splitting threshold. We plot F_1 of the classification under different settings in Figs. 10–12. Moreover, we compute the average increasing rates of F_1 of the class in all scenes as ΔF_1 class to measure the performance improvement between different scale or level configurations. The detailed statistics are listed in Table IV. In Figs. 10–12, the purple lines show the classification quality of the three scenes by using $1s1l$. The classification results of the trees, buildings, and cars are satisfactory when employing the features only extracted from $1s1l$ of the clusters. This means that the LDA can effectively obtain the features from the clusters and reduce the sensitivity to the point density. However, the classification result obtained by using $1s1l$ is the worst in most situations compared with those obtained by using other scales and levels. In Figs. 10–12, from the blue lines $(3s3l)$ versus the red lines $(1s3l)$ and green lines $(3s1l)$ versus the purple lines $(1s1l)$, it can be noted that that use of more scales improves the performance and robustness of the classification on varying point density in general. When more scales are used to classify

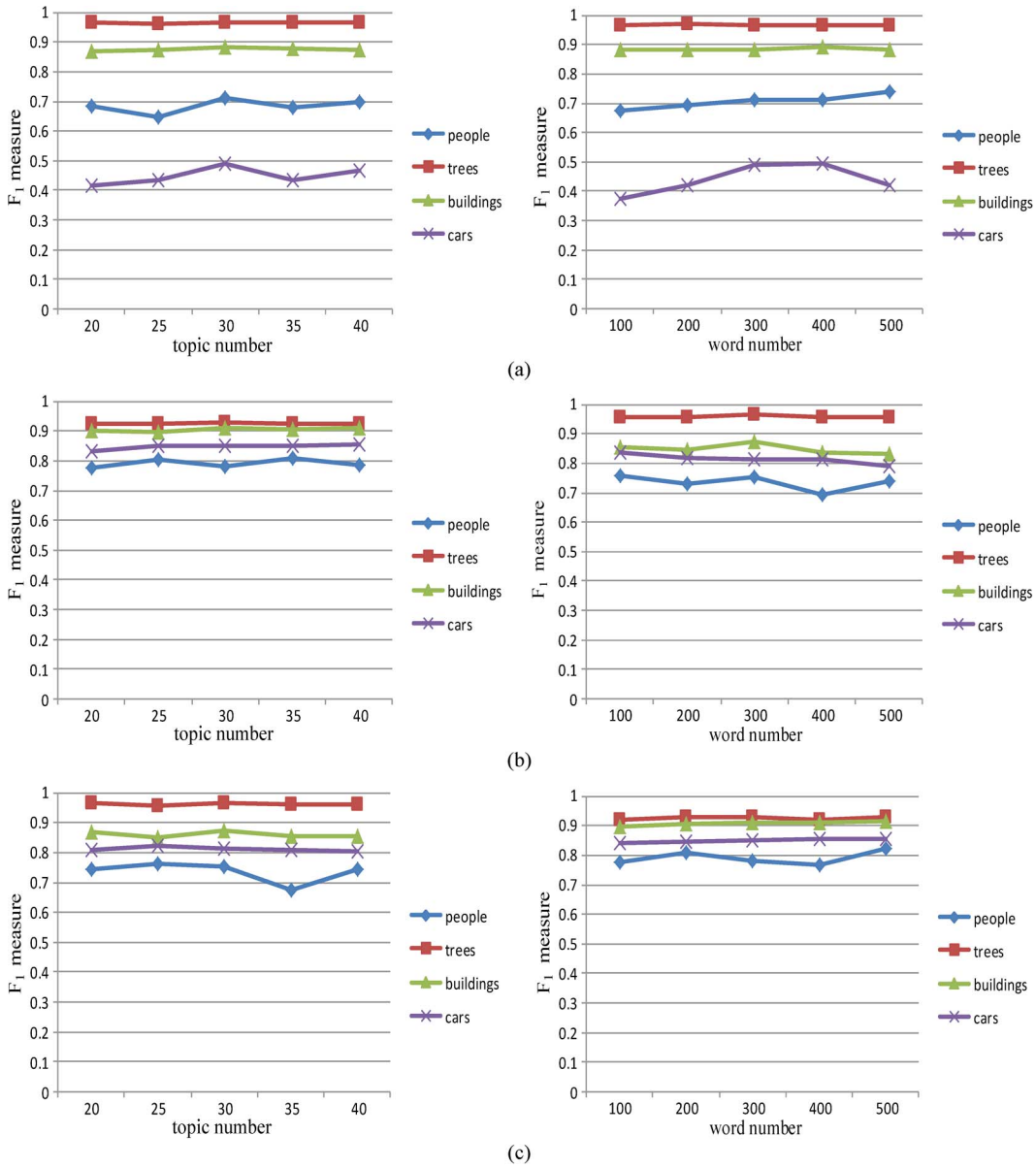


Fig. 9. Impact of the word number and topic number on classification results. (a) Scene I. (b) Scene II. (c) Scene III.

the four classes, the value range of F_1 is smaller than that used with one scale. In some subfigures such as Fig. 10(a) and Fig. 12(a)–(d), F_1 by using one scale is higher than when using three scales. However, the former decreases faster than the latter when the point density decreases, indicating that the explicitly constructed multiscale point clouds of the training process can make the features scale invariant and the classifiers more applicable to different scenes. There is no need to change the training data for the unfamiliar scenes. The hierarchical point clusters improve the performance of the classification for smaller objects, such as cars and people. From the second and third rows in Table IV, the F_1 measure using more levels is improved. Therefore, classifying objects by using the different regions is helpful for improving the classification quality. We also observe that the increasing rate between $3s3l$ and $3s1l$ is higher than that between $1s3l$ and $1s1l$, because when different scales are applied, although the point cluster size of the levels is stable, the region of the point cluster is varying in different scales. Therefore, the classification results obtained by the multiple scales and hierarchies outperform those of all other configurations.

VII. Conclusion and Future Work

In this paper, we present a novel framework based on the MHPCs for effectively extracting the shape features of objects from complex TLS point clouds. It is discriminative and robust for the point clouds with varying point density and data missing. The features extracted from the MHPCs are scale invariant. Moreover, the information about the different levels can be combined during the classification process. The LDA is introduced to effectively extract the features from the MHPCs. It extracts features with the same length from the clusters with different numbers of points. The LDA can maintain the orientation invariance and reduce the sensitivity of the point-based features. The classification results show that our features extracted from

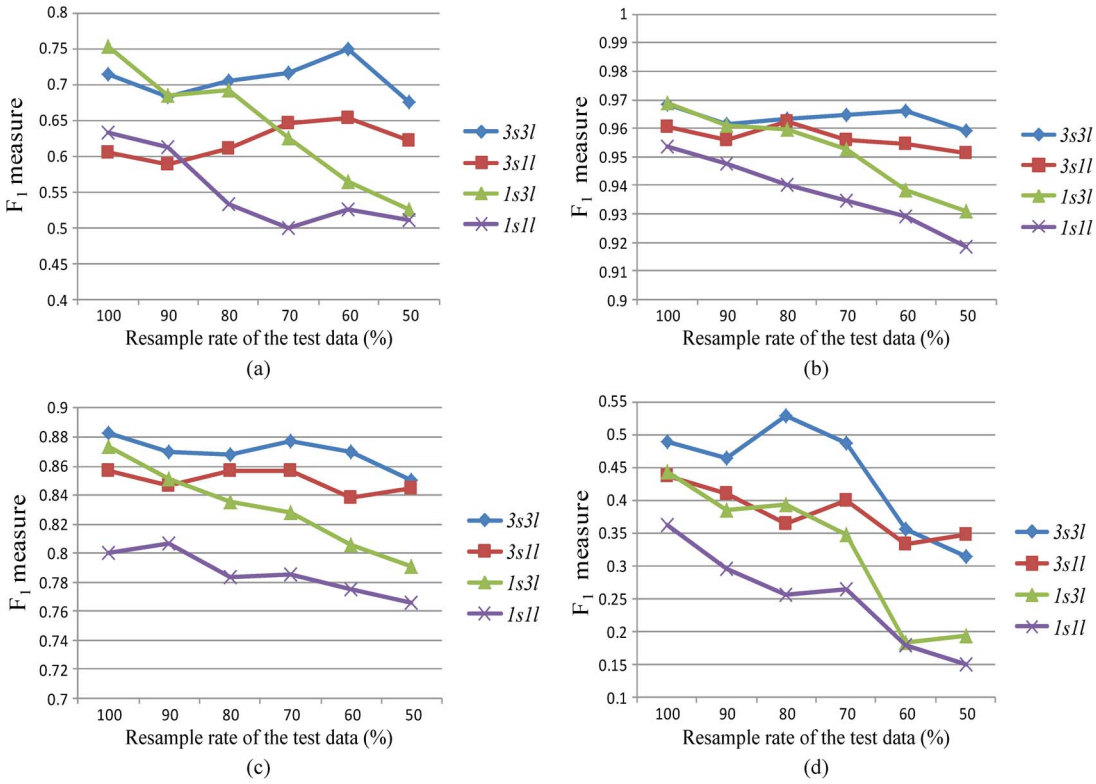


Fig. 10. Classification results of Scene I using the features extracted from the different clusters. (a) People. (b) Trees. (c) Buildings. (d) Cars.

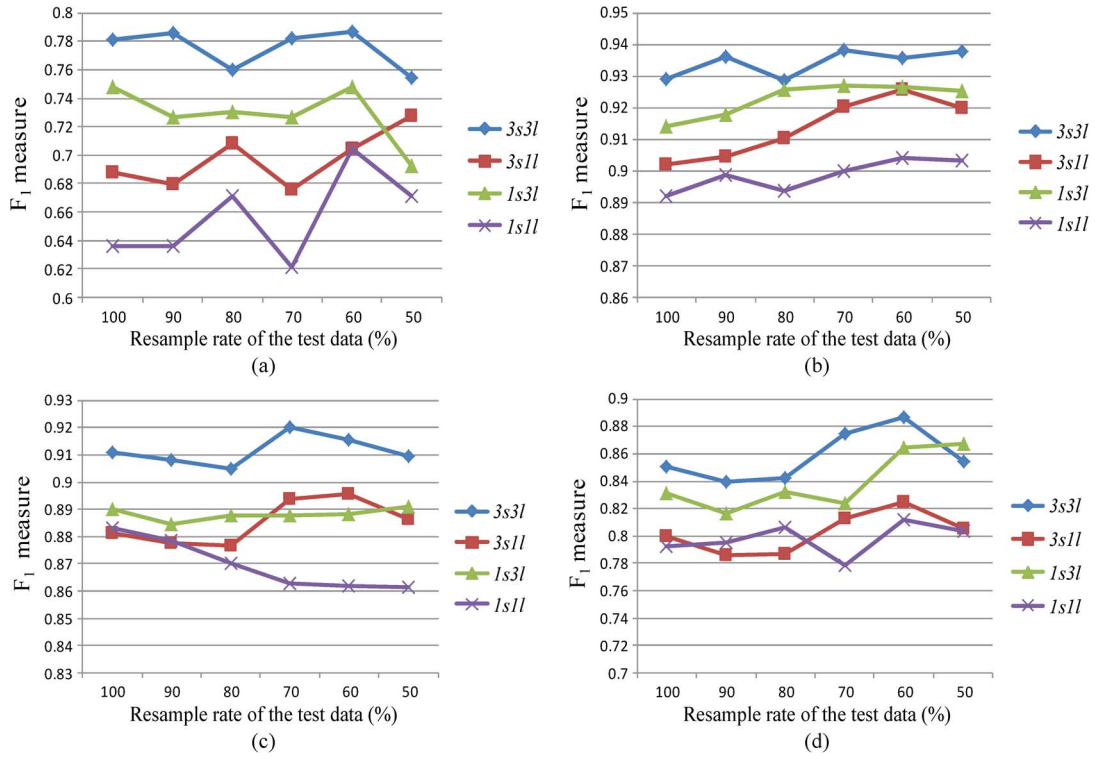


Fig. 11. Classification results of Scene II by using the features extracted from the different clusters. (a) People. (b) Trees. (c) Buildings. (d) Cars.

the MHPCs perform superiorly to the conventional point-based features. Compared with other methods, our approach has a better ability to classify object classes with small size, such as people and cars. The shape features extracted by our approach also can be employed to 3-D shape retrieval [41], [42].

Good clusters and point-based features are helpful to obtain accurate classifications. In future work, we would like to extend this model to incorporate structural information such as the spatial relationships [43] between points and object patches in the classification process. Using models to match the objects of

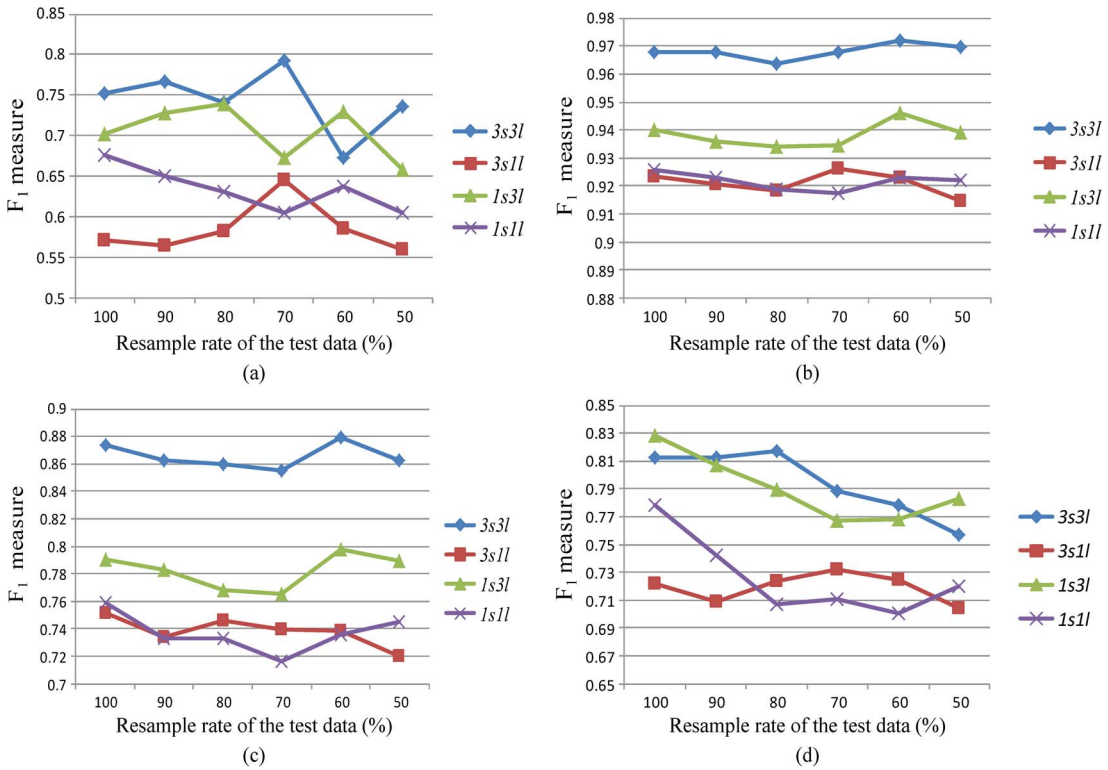


Fig. 12. Classification results of SCENE III by using the features extracted from different clusters. (a) People. (b) Trees. (c) Buildings. (d) Cars.

TABLE IV
COMPARISON OF THE F_1 MEASURE BY USING DIFFERENT SCALES AND LEVELS

	$\Delta F_{1people}$	ΔF_{1trees}	$\Delta F_{1buildings}$	ΔF_{1cars}
3s3l vs. *3s1l*	12.8%	2.7%	5.9%	6.5%
1s3l vs. *1s1l*	8.6%	1.7%	3.7%	5.7%
3s3l vs. *1s1l*	10.8%	3.1%	7.5%	7.3%

point clouds can improve classification accuracy if an effective way is found. We will also use the Graphic Processing Unit (GPU) programming and parallelization scheme to further reduce computing time.

ACKNOWLEDGMENT

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

REFERENCES

[1] S. K. Lodha, E. J. Kreps, D. P. Helmbold, and D. N. Fitzpatrick, "Aerial LiDAR data classification using support vector machines (SVM)," in *Proc. Int. Symp.3D Data Process., Vis. Transmiss.*, Chapel Hill, NC, USA, 2006, pp. 567–574.

[2] S. K. Lodha, D. M. Fitzpatrick, and D. P. Helmbold, "Aerial lidar data classification using AdaBoost," in *Proc. Int. Conf. 3-D Digit. Imag. Model.*, Montreal, QC, Canada, 2007, pp. 435–442.

[3] N. Chehata, L. Guo, and C. Mallet, "Airborne lidar feature selection for urban classification using random forests," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 38, no. 3, pp. 207–212, 2009.

[4] J. Niemeyer, F. Rottensteiner, and U. Soergel, "Conditional random fields for lidar point cloud classification in complex urban areas," *ISPRS Ann. Photogramm., Remote Sens., Spatial Inf. Sci.*, vol. 1–3, pp. 263–268, 2012.

[5] B. K. P. Horn, "Extended Gaussian images," *Proc. IEEE*, vol. 72, no. 12, pp. 1671–1686, Dec. 1984.

[6] S. B. Kang and K. Ikeuchi, "The complex EGI: A new representation for 3-D pose determination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 7, pp. 707–721, Jul. 1993.

[7] M. Hebert, K. Ikeuchi, and H. Delingette, "A spherical representation for recognition of free-form surfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 7, pp. 681–690, Jul. 1995.

[8] M. Weinmann, B. Jutzi, and C. Mallet, "Feature relevance assessment for the semantic interpretation of 3D point cloud data," *ISPRS Ann. Photogramm., Remote Sens., Spatial Inf. Sci.*, vol. II-5/W2, pp. 313–318, 2013.

[9] K. F. West *et al.*, "Context-driven automated target detection in 3D data," in *Proc. Defense Security*, Orlando, FL, USA, 2004, pp. 133–143.

[10] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.

[11] D. Huber, A. Kapuria, R. Donamukkala, and M. Hebert, "Parts-based 3D object classification," in *Proc. IEEE Comp. Soc. Conf. Comput. Vis. Pattern Recog.*, Washington, DC, USA, 2004, vol. 2, pp. II-82–II-89.

[12] A. E. Johnson, O. Carmichael, D. F. Huber, and M. Hebert, "Toward a general 3-D matching engine: Multiple models, complex scenes, efficient data filtering," in *Proc. Image Understanding Workshop*, Monterey, CA, USA, 1998, pp. 1097–1107.

[13] J. J. Caceres and K. C. Slatton, "Improved classification of building infrastructure from airborne lidar data using spin images and fusion with ground-based lidar," in *Proc. Urban Remote Sens. Joint Event*, Paris, France, 2007, pp. 1–7.

[14] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *Proc. Eur. Conf. Comput. Vis.*, Prague, Czech Republic, 2004, pp. 224–237.

[15] 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 Comp. Soc. Conf. Comput. Vis. Pattern Recog.*, New York, NY, USA, 2006, pp. 1605–1614.

[16] 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.

[17] 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, Oct. 2009.
[18] E. H. Lim and D. Suter, "Multi-scale conditional random fields for over-segmented irregular 3D point clouds classification," presented at the IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog., Anchorage, AK, USA, 2008, pp. 1–7.
[19] 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.
[20] 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.
[21] 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.
[22] 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.
[23] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn.*, vol. 42, pp. 177–196, Jun. 2001.
[24] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, May 2003.
[25] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, San Diego, CA, USA, 2005, pp. 524–531.
[26] D. Aldous, "Exchangeability and related topics," *Écoled'Été de Probabilités de Saint-Flour XIII*, vol. 1117, pp. 1–198, 1985.
[27] T. Griffiths, M. Jordan, and J. Tenenbaum, "Hierarchical topic models and the nested Chinese restaurant process," *Adv. Neural Inf. Process. Syst.*, vol. 16, pp. 106–114, 2004.
[28] E. Bart, I. Porteous, P. Perona, and M. Welling, "Unsupervised learning of visual taxonomies," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Anchorage, AK, USA, 2008, pp. 1–8.
[29] J. Sivic, B. C. Russell, A. Zisserman, W. T. Freeman, and A. A. Efros, "Unsupervised discovery of visual object class hierarchies," presented at the IEEE Conf. Comput. Vis. Pattern Recogn., Anchorage, AK, USA, 2008, pp. 1–8.
[30] F. Endres, C. Plagemann, C. Stachniss, and W. Burgard, "Unsupervised discovery of object classes from range data using latent Dirichlet allocation," in *Proc. Robot., Sci. Syst.*, Washington DC, USA, 2009.
[31] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.
[32] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[33] 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.
[34] S. Xu, S. Oude Elberink, and G. Vosselman, "Entities and features for classification of airborne laser scanning data in urban area," *ISPRS Ann. Photogramm., Remote Sens., Spatial Inf. Sci.,* vol. I-4, pp. 257–2662, 2012.
[35] D. M. Blei and J. D. McAuliffe, "Supervised topic models," arXiv preprint arXiv:1003.0783, 2010.
[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] H. Gross and U. Thoennessen, "Extraction of lines from laser point clouds," in *Proc. Photogramm. Image Anal.*, Bonn, 2006, pp. 87–91.
[39] J. Thongkam, G. Xu, and Y. Zhang, "AdaBoost algorithm with random forests for predicting breast cancer survivability," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Hong Kong, 2008, pp. 3062–3069.
[40] J. C.-W. Chan and D. Paelinckx, "Evaluation of random forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2999–3011, Jun. 2008.
[41] M. Zhang, L. Zhang, P. T. Mathiopoulos, Y. Ding, and H. Wang, "Perception-based shape retrieval for 3D building models," *ISPRS J. Photogramm. Remote Sens.*, vol. 75, pp. 76–91, Jan. 2013.
[42] J. Chen, C. Lin, P. Hsu, and C. Chen, "Point cloud encoding for 3D building model retrieval," *IEEE Trans. Multimedia*, vol. 16, no. 2, pp. 337–345, Feb. 2014.
[43] Z. Wang *et al.*, "A structure-aware global optimization method for reconstructing 3-D tree models from terrestrial laser scanning data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5653–5669, Sep. 2014.

Zhen Wang is currently working toward the Ph.D. degree in geography with the Beijing Normal University, Beijing, China.

His research interests include the application of light detection and ranging in classification of the ground objects and 3-D urban modeling.

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

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

Tian Fang (M'13) received the bachelor's and master's degrees in computer science and engineering from the South China University of Technology, Guangzhou, China, in 2003 and 2006, respectively, and the Ph.D. degree in computer science and engineering from the Hong Kong University of Science and Technology (HKUST), Kowloon, Hong Kong, in 2011.

He is currently a Postdoctoral Researcher at HKUST, where he now works on projects related to real-time 3-D reconstruction and recognition. His research interests include image-based modeling, image segmentation, recognition, and photorealistic rendering.

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, University of British Columbia, Vancouver, BC, Canada, where he was a Faculty Member until 2003, holding the rank of Professor since 2000. He is currently the Director of Research at the Institute for Astronomy Astrophysics Space Applications and Remote Sensing, National Observatory of Athens (ISARS/NOA), Athens, Greece, and a Professor in digital communication at the Department of Informatics and Telecommunication, National and Kapodistrian University of Athens, Athens. For the last 25 years, he has been conducting research mainly on the physical layer of digital communication systems for terrestrial and satellite applications. He is also interested in channel characterization and measurements, modulation and coding techniques, Single-Input Multi-Output/Multiple-Input Single-Output, Ultra Wide Band, Orthogonal Frequency Division Multiplexing, software/cognitive radios, and green communications. He has also research activities in the fields of remote sensing and photogrammetry.

Dr. Mathiopoulos has been or currently serves on the editorial board of several archival journals, including the *The Institution of Engineering and Technology Communications* and the IEEE TRANSACTIONS ON COMMUNICATIONS in 1993–2005. Since 1993, he has served on a regular basis as a Scientific Advisor and a Technical Expert for the European Commission (EC). In addition, since 2001, he has served as the Greek representative to high-level committees in the EC and the European Space Agency. He has been a member of the Technical Professional Conference of more than 50 international conferences, a TPC Vice Chair for the 2006-S IEEE Vehicular Technology Conference and 2008-F IEEE VTC, and a Cochair of FITCE2011. He has delivered numerous invited presentations, including plenary lectures, and he has taught many short courses all over the world. He was an African Scientific Institute Fellow, a Killam Research Fellow, and a corecipient of two best paper awards.

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 at The Hong Kong Polytechnic University, Kowloon, Hong Kong. Between 2008 and 2009, he was a Visiting Scholar at 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 at Tongji University. He authored more than 40 publications in international journals. His current research interests include remote sensing, Geographic Information System, trust in spatial data, 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. Moreover, he was a recipient of the State Natural Science Award (Second place) from the State Council of the Peoples' Republic of China in 2007 and the National Natural Science Funds for Distinguished Young Scholar in 2013.

Huamin Qu (M'07) received the B.S. degree in mathematics from the Xi'an Jiaotong University, Xi'an, China, and the M.S. and Ph.D. degrees in computer science from the Stony Brook University, Stony Brook, NY, USA, in 2004.

He is an Associate Professor with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong. He has coauthored more than 60 refereed papers, including 20 papers in the IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS (TVCG). His main research interests are in visualization and computer graphics.

Prof. Qu serves on the steering committee of the IEEE Pacific Visualization Conferences and is an Associate Editor of IEEE TVCG. He received the Honorable Mention for the Best Paper Award at IEEE Visualization 2009 and is a recipient of the 2009 IBM Faculty Award.

Zhiqiang Xiao received the Ph.D. degree in geophysical prospecting and information technology from Central South University, Changsha, China, in 2004.

From 2004 to 2006, he was a Postdoctoral Research Associate with Beijing Normal University, Beijing, China, where he is currently with the State Key Laboratory of Remote Sensing Science. His research focuses on retrieval of land biophysical parameters from remotely sensed data, assimilating radiometric observations into dynamic models.

Fang Li is currently working toward the master's degree with the School of Geography, Beijing Normal University, Beijing, China.

Her research interests include location-based service and remote sensing imagery processing.

Dong Chen received the Ph.D. degree in geographical information sciences from Beijing Normal University, Beijing, China, in 2013.

He is currently an Assistant Professor with the Nanjing Forestry University, Nanjing, China. His research interests include image- and Light Detection and Ranging-based segmentation and reconstruction, full-waveform LiDAR data processing, and other remote sensing applications in the field of forest ecosystems.