A Hierarchical Approach for Point Cloud Classification With 3D Contextual Features

Chen-Chieh Feng and Zhou Guo

***Abstract*—Classifying point cloud of urban landscapes plays essential roles in many urban applications. However, automating such a task is challenging due to irregular point distribution and complex urban scenes. Incorporating contextual information is crucial in improving classification accuracy of point clouds. In this article, we propose a hierarchical approach for point cloud classification with 3-D contextual features, which comprises three steps: segment-based classification with primitive features and a random forest classifier; extracting novel 3-D contextual features from the initial labels considering spatial relationships between neighboring segments and semantic dependencies; and refining classification with a combination of primitive features and spatial contextual features, and a hierarchical multilayer perceptron classifier that considers primitive features and spatial contextual features at different levels. The proposed method was tested on two point cloud datasets: the National University of Singapore (NUS) dataset and the Vaihingen benchmark dataset of the International Society of Photogrammetry and Remote Sensing. The evaluation results showed that the proposed method achieved an overall accuracy of 92.51% and 82.34% for the NUS dataset and Vaihingen dataset, respectively. The feature importance evaluation showed that 3-D spatial contextual features contributed useful information for discriminating different classes, such as roof, facade, grassland, tree, and ground. Quantitative comparisons further showed that the proposed method is more advantageous, especially in the detection of class roof and facade.**

***Index Terms*—Classification, contextual feature, hierarchical classifier, point cloud.**

I. Introduction

THE drive toward smart cities around the world has necessitated the development of 3-D spatial data infrastructures given that they provide precise descriptions of the man-made structures and representations of natural resources in 3-D, and quantifiable pieces of evidence into urban dynamics when armed with Internet-of-Things and social media feeds. Point cloud data generated by light detection and ranging (LiDAR) is one of the most popular methods to develop such 3-D spatial infrastructures. LiDAR point cloud data have been utilized in many applications and research topics, e.g., creation of large-scale city models [1], generation of digital terrain model [2], mapping of vegetation [3], reconstruction of 3-D buildings [4], and detection of road markings [5] or changes [6]. Most of these applications require point cloud classification as a basic LiDAR processing step, which is to assign each point a semantic label such as ground and grassland [7]. Due to the enormous number of points in the point cloud data, automatic classification approaches are often adopted. However, the automation of point cloud classification in urban areas is challenging because of the complexity of urban scenes and a high level of heterogeneity.

Existing LiDAR classification approaches can be categorized into point- or segment-based [8]. For both types of approaches, various standard supervised methods have been adopted, such as decision trees [9], random forests [10], support vector machines [11], Gaussian Mixture Models [12], AdaBoost [13], and Bayesian discriminant classifiers [14]. Two important factors affecting these standard supervised classification are representative samples and discriminant features. The former, despite its importance, received minor attention because most published studies verify their performances using benchmark point cloud where labels of training data are often available. An exception is the article in Feng and Guo [15] where an automatic method for sample selection with the aid of 2-D land cover maps and a built topological graph is provided. For the latter, spectral, geometrical, and eigen-based features are commonly used [7], [15]. Both point- and segment-based approaches extract eigen-based features from a 3-D covariance matrix, which comprises 3-D coordinates of a cluster of points, to represent local characteristics of points. The discriminant features are chosen either manually or through a variety of feature selection approaches [16]–[18]. However, these discriminant features often result in unsatisfactory classification results for complex urban scenes because each classified entity (point or segment) is treated independently, i.e., without considering labels of its neighbors or contextual information. Such deficiency has led to point cloud classification approaches to incorporate contextual information by either adding a smooth constraint in a probabilistic graphical model, adopting contextual information in a post-processing step, or defining contextual features for classification.

For the first approach, i.e., adding a smooth constraint in a probabilistic graphical model, the most commonly used probabilistic graphical models are Markov random fields (MRF) and condition random fields (CRF). In both models, spatial contextual information is expressed by the pairwise potentials in a posterior energy function, and penalties are imposed on

Manuscript received December 24, 2020; revised March 1, 2021 and April 8, 2021; accepted April 17, 2021. Date of publication May 4, 2021; date of current version May 26, 2021. This work was supported by Singapore Ministry of Education Academic Research Fund Tier 1 FY2019-FRC4-008 and the Fundamental Research Funds for the Central Universities, Sun Yat-sen University. *(Corresponding author: Zhou Guo.)*

Chen-Chieh Feng is with the Department of Geography, National University of Singapore, 117570, Singapore (e-mail: geofcc@nus.edu.sg).

Zhou Guo is with the School of Geospatial Engineering and Science, Sun Yat-Sen University, Zhuhai 519082, China (e-mail: guozhou@pku.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3077568

neighboring entities (points or segments) with different labels [19], [20]. Munoz *et al.* [21] applied associate Markov network (AMN) and adopted a functional gradient approach to learn high-dimensional parameters of random fields. The shortcoming of AMNs is that the pairwise potential function is constant (i.e., NULL) when a pair of entities have different labels, which is too rigorous in many cases and often results in over smoothing and error propagation. To overcome this shortcoming, non-AMNs (non-AMNs) are adopted. Shapovalov *et al*. [22] proposed a non-AMNs technique with three contextual features for the pairwise potentials for laser scanning data classification and showed that non-AMN models performed better than AMN models. CRF is a more general model for considering the context in classification and providing a flexible statistical framework [23]. It differs from MRF in how potentials are modeled. Laible *et al.* [24] adopted a CRF-based model for classifying fused 3-D LiDAR and camera data and showed that CRF achieved better performance than an MRF-based approach. For the pairwise potentials of CRF, arbitrary classifiers can be used, and the earlier research efforts tended to adopt simple models, such as Potts model [25], the linear model of Generalized linear model [26], and the random forest classifier [20]. Despite the simplicity and better performance than the MRF-based approaches, the probabilistic graphical models in CRF oversimplify spatial contextual information in two respects. First, the spatial relationship considered is restricted to neighboring relationship. Distance and directional relationships, which can be equally important for point cloud classification, are not well captured. Second, these models are semantic independent regarding the spatial relations considered because they only concern whether two neighboring entities have the same labels, ignoring the possibility that the spatial relationships between different pairs of entities vary greatly and are highly dependent on the labels of entity pairs.

The second approach, i.e., adopting contextual information in the postclassification step, involves first obtaining initial labeling through a pointwise probabilistic classification, and second improving the initial labeling with contextual information. They can normally be divided into local or global optimization based on the dependent range of neighboring points. Local optimization methods generally apply filters in the local neighborhood. For example, Li *et al*. [27] optimized the initial labeling by a local optimization approach of decision tree based on weak priors. Although the initial classification results were greatly improved by the local optimization, the rules of weak priors including thresholds are handcrafted by experience, which can be subjective and lack extensibility to other environments, let alone the challenges involving the identification of thresholds for complex scenes. Global optimization methods find the solution for the lowest energy cost in a global context. Based on global optimization, Landrieu *et al*. [28] proposed a structured regularization framework for spatially smoothing semantic labeling of 3-D point clouds. Similar work was conducted in Huang *et al.* [29], where a global graph-based optimization based on MRF was performed to optimize the initial classification results obtained using embedded deep features. Li *et al*. [30] proposed a probabilistic label relaxation (PLR) approach, which includes two stages. In the first stage, the optimal local neighborhood was estimated, thus collecting neighboring point information for each point to identify initial label probabilities. In the second stage, the initial label probabilities were iteratively enhanced with PLR by incorporating spatial contextual information and contextual constraints. The two stage process was able to detect most wrongly labeled regions in the initial classification and correct the misclassification to improve significantly the overall classification accuracy. The main limitation of global optimization is that it may not work very well with points of incomplete objects or occluded objects.

The last approach, i.e., defining contextual features for classification, differs from the above two types of methods in abstracting the contextual information into features and subsequently employing them in point cloud classification. It was first used in 2-D image classification [31] and later extended to point clouds [32] where the contextual feature of the distance between objects and the nearest street was extracted and OpenStreetMap was incorporated to indicate the existence of roads. Instead of incorporating other data sources, Yang *et al*. [33] first extracted road surfaces from point clouds, and subsequently segmented the remaining points into individual candidate objects. For each candidate object, three types of contextual features–relative position, relative direction, and spatial distribution pattern–were then defined and calculated based on the extracted road surfaces. The contextual features, along with other features such as geometric features, were fed into an SVM classifier to label the candidate objects. While useful, the efficacy of such contextual features highly depends on the precision of the reference objects, i.e., the extracted road surfaces.

More recently, deep learning approaches have increasingly been adopted to extract features suited for 3-D point cloud classification. Zhou and Tuzel [34] divided the point clouds into regular voxels and encoded each voxel via voxel feature encoding layers based on the statistical attributes of contained points. Qi *et al*. [35] proposed PointNet, which operates directly on raw points but is limited to capturing local structures. PointNet++ [36] extended PointNet by learning local structures at multiple scales. Although deep learning techniques can also capture contextual information with hierarchical and multi-scale convolutional layers, its black box nature makes it challenging to account for the spatial relations between different semantic objects with deep features.

This article aims to improve point cloud classification results by incorporating 3-D spatial contextual information. It does so by using novel contextual features, which incorporate 3-D spatial relationships in addition to the commonly used neighborhood relationship, and at the same time leverage semantic dependencies between neighboring objects. To obtain semantic information, initial classification of point clouds with primitive (non-contextual) features and a standard supervised classifier is conducted. In this article, primitive features refer to features that are extracted from single objects without considering spatial relationships between two or multiple objects, while spatial contextual features are features derived from spatial relationships of neighboring objects with the initial labels. Further classification is conducted by combining the newly extracted spatial contextual features and the primitive features to refine the

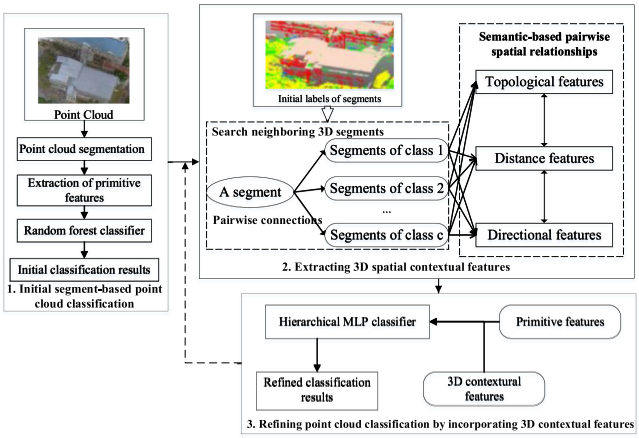

Fig. 1. Framework of the proposed approach for point cloud classification with 3-D contextual features.

initial classification results. Recognizing that contextual features are high-level features and should be treated separately from primitive features for classification, we develop a hierarchical classifier that handles features at different levels. Our main contributions are as follows.

1) We propose novel 3-D spatial contextual features, which consider, on top of neighborhood relationships, the additional spatial relationships of topology, distance, direction, and semantic dependencies. The proposed 3-D spatial contextual features are extracted based on initial classification results, which can be improved iteratively by incorporating spatial contextual features and refined classification. Additionally, the reference objects for the 3-D contextual features can be associated with any category instead of one specific category considered in existing research efforts, so that they can apply to scenes where reference objects are unknown or missing.
2) We develop a hierarchical classifier, which separates primitive features and spatial contextual features to different levels for point cloud classification. The evaluation conducted in this article shows that the hierarchical classifier achieves a higher overall accuracy (OA) of point cloud classification than the nonhierarchical counterpart.

The remainder of this article is organized as follows. In Section II, we introduce our method in detail. The performance evaluation is provided in Section III. Section IV discusses the experiment and Section V summarizes with conclusions and future work.

II. Hierarchical Point Cloud Classification With Spatial Contextual Features

Fig. 1 illustrates the overall framework of the proposed method. It consists of three main steps.

1) Initial segment-based point cloud classification.
2) 3-D spatial contextual feature extraction.
3) Classification refinement with 3-D contextual features and a hierarchical multilayer perceptron (H-MLP) network.

A. Initial Segment-Based Point Cloud Classification

This step aims to provide the initial labeling of the input 3-D point cloud. It adopts a segment-based method due to its advantages in contextual classification. The classification involves three components: point cloud segmentation, extraction of primitive features, and random forest classification.

1) Point Cloud Segmentation: Segmentation groups the neighboring points with the same labels and divides the point cloud into meaningful and nonoverlapping subsets, i.e., segments. In this article, the segmentation process is conducted by region growing from seed points to neighboring points. For the selection of seed points, first, the curvature value of each point is computed using the method of [48]. The curvature values are then sorted and the points with minimum curvature are chosen and placed in a seed list. Second, the seed list is inserted with neighboring points of the current point before the current point is removed. For the insertion process, a similarity index based on the homogeneity of geometric characteristics is computed. It terminates when no neighboring points meeting the homogeneity criteria are found. For the definition of neighborhood for points, this article adopts k-nearest-neighborhood (KNN), which is based on a fixed number of nearest points and approximates the density-adaptive search in an unevenly distributed point cloud, because point clouds are characterized by varying point densities. To identify the optimal number of nearest points k, the two-step approach proposed by [15] is employed. In the first step, a series of segmentation with different k values are conducted. In the second step, segmentation results are quantitatively evaluated using degree of oversegmentation and degree of undersegmentation. The *k* value corresponding to the optimal segmentation performance, i.e., the lowest oversegmentation and undersegmentation, is then selected.

2) Extraction of Primitive Features: Features are extracted based on the 3-D segments generated in the previous section. For colored point clouds, a total of 24 segment features, which are either eigen-based, spectral, or geometrical features (see Table I), are extracted. For eigen-based features, λ_1, λ_2, and λ_3 are the eigenvalues of a point segment ($\lambda_1 > \lambda_2 > \lambda_3$) and computed by a covariance matrix comprised of 3-D coordinates of points within the segments. n_x, n_y, and n_z are values of the normal vector of a segment, which corresponds to the eigenvector of the smallest eigenvalue λ_3. The other eigen-based features, including the sum, omnivariance, eigenentropy, anisotropy, planarity, linearity, surface variation, and sphericity, are arithmetic combinations of λ_1, λ_2, and λ_3. The spectral features are specific for colored point clouds. Mean and standard deviational values of red, green, and blue bands are computed. For geometrical features, they include the number of points within a segment, projective area of the segment on its best-fitted surface plane, local point density of a segment defined by dividing the number of points with the projective area of the segment, and average height of a segment. For non-colored point clouds, the only color information is intensity, and the segment-based spectral features are mean and standard deviational values of intensity, thus reducing the total number of primitive features to 20.

TABLE I
THE PRIMITIVE FEATURE SET BASED ON 3-D POINT SEGMENTS

Eigen-based features	F_1	Eigenvalue_1	λ_1
	F_2	Eigenvalue_2	λ_2
	F_3	Eigenvalue_3	λ_3
	F_4	Normal_x	n_x
	F_5	Normal_y	n_y
	F_6	Normal_z	n_z
	F_7	Sum	$\lambda_1+\lambda_2+\lambda_3$
	F_8	Omnivariance	$\sqrt[3]{\lambda_1 * \lambda_2 * \lambda_3}$
	F_9	Eigenentropy	$-\sum_{i=1}^{3} \lambda_i * Ln(\lambda_i)$
	F_{10}	Anisotropy	$(\lambda_1 - \lambda_3)/\lambda_1$
	F_{11}	Planarity	$(\lambda_2 - \lambda_3)/\lambda_1$
	F_{12}	Linearity	$(\lambda_1 - \lambda_2)/\lambda_1$
	F_{13}	Surface Variation	$\lambda_3/(\lambda_1 + \lambda_2 + \lambda_3)$
	F_{14}	Sphericity	λ_3/λ_1
Spectral features	F_{15}	Mean Red	M_R
	F_{16}	Mean Green	M_G
	F_{17}	Mean Blue	M_B
	F_{18}	Red Deviation	Std_R
	F_{19}	Green Deviation	Std_G
	F_{20}	Blue Deviation	Std_B
Geometrical features	F_{21}	Number of points	#N
	F_{22}	Area	S
	F_{23}	Local point density	LPD
	F_{24}	Average height	H_z

B. 3-D Spatial Contextual Feature Extraction

As the primitive features are extracted from individual segments, the spatial relationships between different classes in a 3-D environment are not reflected in the initial classification. To address this limitation, new 3-D contextual features are defined and extracted. The novelties of these 3-D contextual features are reflected in two aspects: multiple spatial relationships including topology, distance and direction, instead of the simple neighboring relationship adopted in most of existing contextual classification approaches of point clouds, are taken into account; and the proposed 3-D contextual features are semantic-dependent, i.e., the pairwise intersections take semantic labels of two neighboring segments into account. For each type of spatial relationship and each pair of semantic labels, a 3-D spatial contextual feature is developed. Below the definition of neighborhood between 3-D candidate segments will be introduced first, followed by the definition of 3-D contextual features in terms of different spatial relationships.

1) Definition of Neighborhood for Candidate Segments: Existing literature has suggested at least three approaches to define neighborhood of 3-D point clouds through adjacency. The first approach describes the neighborhood between individual 3-D points using either a sphere, a cylinder, or KNN [37]. The second approach is by supervoxel in 3-D [8], [38]. The neighborhood of a supervoxel is identified through adjacency, which is unambiguous as supervoxels have regular shapes. The third approach to define the neighborhood of point clouds is through 3-D point segments. The neighboring relationship defined this way is, however, vague and thus seldomly been used because segments are neither organized in regular shapes nor share the same contours as 2-D image segments do. In this article, we adopt cylindrical neighborhood but extend it from single points to segments by the following steps. First, for each point, its cylindrical neighboring points, i.e., those points who are within a radius search *r* of the center point on the *XOY* plane, are identified and stored. Second, two point segments are neighbors if at least one point each from the two segments are neighbors.

2) Definition of 3-D Contextual Features: The 3-D contextual features of a segment are defined based on its neighboring segments and are related to their semantic labels. As stated above, our 3-D spatial contextual features consider multiple spatial relationships, including topology, distance, and direction, and are semantic dependent. The number of defined 3-D contextual features for each segment is thus $N*C$, where N is the number of metrics measuring spatial relationships, and C the number of semantic classes. Each contextual feature measures a spatial relationship with one class. Below we explain how to define 3-D contextual features in terms of three types of spatial relationships: topology; distance; and direction.

To capture topological relationships while considering semantics, the degree of overlap (DoV) is defined. Let S_i and S_j be two neighboring point segments and S_i is the target segment, the DoV between S_i and S_j is defined as

$$\mathrm{DoV}\ (S_i, S_j) = \frac{\mathrm{Area}_{XOY}\ (S_i \cap S_j)}{\mathrm{Area}_{XOY}\ (S_i)} \tag{1}$$

where $\mathrm{Area}_{XOY}(S_i \cap S_j)$ is the overlapping area of S_i and S_j on the *XOY* plane, $\mathrm{Area}_{XOY}(S_i)$ is the projective area of S_i on the *XOY* plane. Assume there are M neighboring segments of S_i with semantic label c, the measure of DoV for segment S_i in terms of class c is defined as the total DoV values of the M segments

$$\mathrm{Dov}_i^c = \sum_{j=1}^{M} \mathrm{DoV}\ (S_i, S_j)\,. \tag{2}$$

For distance, two metrics measuring the horizontal and vertical distances are defined. Let N_i and N_j be the number of points in segment S_i and S_j, and S_i^r and S_j^t are the rth and tth point, horizontal and vertical distances between S_i and S_j are represented by $\mathrm{Dis}_H(S_i, S_j)$ and $\mathrm{Dis}_z(S_i, S_j)$ and defined as

$$\mathrm{Dis}_H\ (S_i, S_j) = \frac{\sum_{r=1}^{N_i}\sum_{t=1}^{N_j}\sqrt{\left(S_i^r\,(x) - S_j^t\,(x)\right)^2 + \left(S_i^r\,(y) - S_j^t\,(y)\right)^2}}{N_i * N_j} \tag{3}$$

$$\mathrm{Dis}_z\ (S_i, S_j) = \frac{\sum_{r=1}^{N_i}\sum_{t=1}^{N_j}\left(S_j^t\,(z) - S_i^r\,(z)\right)}{N_i * N_j} \tag{4}$$

where $S_i^r(x)$, $S_i^r(y)$, and $S_i^r(z)$ are the x, y, and z coordinates of point S_i^r individually. Note that the vertical distance can be negative when segment S_i is under segment S_j. Similarly,

measures of horizontal and vertical distance for segment S_i in terms of class c is defined as the average Dis_H and Dis_z values of the M neighboring segments with label of c

$$\mathrm{Dis\_}H_i^c = \frac{\sum_{j=1}^{M} \mathrm{Dis}_H\left(S_i, S_j\right)}{M} \tag{5}$$

$$\mathrm{Dis\_}Z_i^c = \frac{\sum_{j=1}^{M} \mathrm{Dis}_Z\left(S_i, S_j\right)}{M}. \tag{6}$$

For direction, two metrics measuring the horizontal and vertical direction are also defined. The horizontal direction of two 3-D point segments is measured based on the direction of their projective polygons on the *XOY* plane. The vertical direction of two 3-D point segments is measured based on the direction of their normal vectors. The horizontal and vertical direction between S_i and S_j are represented by $\mathrm{Dir}_H(S_i, S_j)$ and $\mathrm{Dir}_z(S_i, S_j)$ and defined as

$$\mathrm{Dir}_H\ \left(S_i, S_j\right) = \begin{cases} \tan^{-1} \frac{\overline{S_j}(y) - \overline{S_i}(y)}{\overline{S_j}(x) - \overline{S_i}(x)} & \overline{S_j}\left(x\right) \neq \overline{S_i}\left(x\right) \\ \frac{\pi}{2} & \overline{S_j}\left(x\right) = \overline{S_i}\left(x\right) \text{ and } \overline{S_j}\left(y\right) > \overline{S_i}\left(y\right) \\ -\frac{\pi}{2} & \overline{S_j}\left(x\right) = \overline{S_i}\left(x\right) \text{ and } \overline{S_j}\left(y\right) < \overline{S_i}\left(y\right) \end{cases} \tag{7}$$

$$\mathrm{Dir}_z\ \left(S_i, S_j\right) = \cos^{-1} \frac{\left|\overrightarrow{N_{S_i}} . \overrightarrow{N_{S_j}}\right|}{\left|\overrightarrow{N_{S_i}}\right| \left|\overrightarrow{N_{S_j}}\right|} \tag{8}$$

where $\overline{S_i}$ and $\overline{S_j}$ are the center points of segment S_i and S_j on the *XOY* plane, $\overrightarrow{N_{S_i}}$ and $\overrightarrow{N_{S_j}}$ the normal vectors of S_i and S_j. Similarly, the measure of horizontal and vertical direction for segment S_i in terms of class c is defined as the average Dis_H and Dis_z values of the M neighboring segments with the label of c

$$\mathrm{Dir\_}H_i^c = \frac{\sum_{j=1}^{M} \mathrm{Dir}_H\left(S_i, S_j\right)}{M} \tag{9}$$

$$\mathrm{Dir\_}Z_i^c = \frac{\sum_{j=1}^{M} \mathrm{Dir}_Z\left(S_i, S_j\right)}{M}. \tag{10}$$

In summary, for a point cloud with n different classes, a total of $5 * n$ 3-D contextual features are computed for each segment.

C. Classification Refinement With 3-D Contextual Features and a Hierarchical-MLP Network

The set of 3-D contextual features, once developed, are combined with primitive segment features as input to refine the initial classification results. Supervised classification is adopted for this purpose. According to Aksoy *et al.* [46], features derived from single objects belong to basic level of information, while features derived from spatial relationships between two or multiple objects are considered as a higher representation of spatial information. The 3-D spatial contextual features are deemed higher level features than the primitive segment features. In addition, Qiao *et al.* [47] pointed out that typically, lower level features are employed in the early classification stages for finding basic landscape information, while spatial relationships are used in the later stages for determining the final classes.

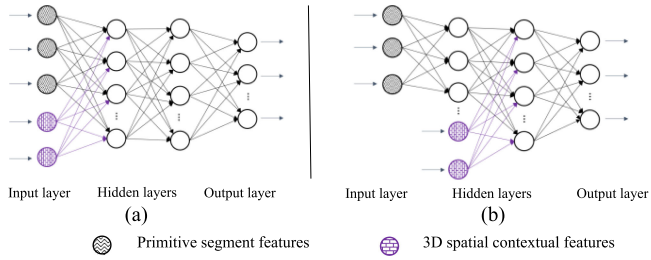

Fig. 2. Structure of (a) MLP network and (b) H-MLP network.

Standard supervised classifiers that treat all features equally are not suitable to handle such cases. To address this issue, we build a H-MLP network, which is an improvement of MLP network with the ability to handle different levels of features for both training and classification.

Fig. 2(a) and (b) shows the traditional MLP and the proposed H-MLP networks. Both networks contain one input layer, one output layer, and two hidden layers. The H-MLP network differs from the traditional MLP network in its input layer and the first hidden layer. In an H-MLP network, the input features are divided into primitive features and 3-D contextual features. The input layer imports only the primitive features, with each feature connecting to a neuron in the input layer. In the first hidden layer, the 3-D contextual features are incorporated as individual neurons along with all the primitive features, also as individual neurons that are fully connected to the neurons in the input layer. Aside from the input layer, all neurons in other layers are fully connected to neurons in the next layer.

Assume $X_0 = \ (X_t,\ X_s)$ is the input feature vector, in which X_t represents primitive features defined in Section II-B-2 and X_s represents 3-D contextual features defined in Section II-C-2. For MLP network, the output of the first hidden layer is $f(W_1 X_0 + b_1)$, where W_1 and b_1 are the weight matrices and the bias term of the first hidden layer in the MLP network, and $f(x)$ is an activation function. For H-MLP network, the output of the first hidden layer is $[f(W_1' X_t + b_1'), X_s]$, which concatenates the transformation of primitive features and the higher level 3-D contextual features into one feature vector. W_1' and b_1' are the weight matrices and bias term of the first hidden layer in the H-MLP network. The output of the $(i-1)$th hidden layer $(i > 1)$ is linked to the input of the ith hidden layer, whose output is $X_i = \ f(W_i' X_{i-1} + b_i')$. To sum up, the output of the ith hidden layer can be represented as:

$$X_i = \begin{cases} \left[f\left(W_i^{'} X_t + b_i^{'}\right), X_s\right] \ i = 1 \\ f\left(W_i^{'} X_{i-1} + b_i^{'}\right) \ \ i > 1 \end{cases} \tag{11}$$

where X_i is the output of the i-th hidden layer, W_i' and b_i' the weight matrices and bias term of the ith hidden layer in the H-MLP network. In this study, $f(x)$ is set as the rectified linear unit function, which is a commonly used activation function in a neural network. The number of hidden layers is at least two in the H-MLP network.

Assume that there are k $(k > 1)$ hidden layers in the H-MLP network, the output of the last hidden layer should be X_k, which forms the input to the network's output layer. The output of the

Fig. 3. NUS dataset.

output layer is connected to $\mathrm{softmax}(W'_{k+1}X_k + b'_{k+1})$, where $\mathrm{softmax}(x)$ is a generalization of logistic regression function that can be used for multilabel classification. Each neuron in the output layer of H-MLP is linked with a semantic class, therefore, the number of neurons in the output layer equals to that of semantic classes to be classified. Let $\tau = \{\tau^i\}_{i=1}^c$ be the set of semantic classes with c being the total number of classes. For a point segment j, assume $\delta_j = W'_{k+1}X_k + b'_{k+1}$ is the vector input to the $\mathrm{softmax}$ function, with δ_j^i linking to the ith neuron in the output layer, the probability of segment j belonging to class τ^i is denoted by $p(y = \tau^i \,|\delta_j)$, which can be modeled by the $softmax$ function:

$$p\left(y = \tau^i \,\delta_j\right) = \frac{e^{\delta_j^i}}{\sum_{i=1}^{c} e^{\delta_j^i}}. \tag{12}$$

Segment j is assigned the predicted label of τ^i with the highest probability, i.e., $y_j = \operatorname{argmax} p(\tau^i)$.

To train the H-MLP network, a large number of sample segments with reference labels are needed. The predicted labels are compared with the reference labels, and the level of misclassification is evaluated with the cross-entropy loss function

$$L = -\sum_{i=1}^{N}\sum_{r=1}^{c} y'_{i,r}\log\left(p\left(y_i = \tau^r\right)\right) \tag{13}$$

where N is the number of point segments, c the number of classes, $y'_{i,r}$ the reference label value (0 or 1) of the ith segment in terms of class τ^r, $p(y_i = \tau^r)$ the computed probability that the ith segment belongs to class τ^r. Once the loss function is defined, parameters in the H-MLP network including the weight matrices W' and the bias vector b' are to be learned. These parameters are optimized by minimizing the cross-entropy loss with the stochastic gradient descent algorithm. The network is trained iteratively until the stopping criterion is met.

III. Experiment Analysis

A. Study Area and Datasets

Two datasets are adopted to verify the effectiveness of the proposed approach. The first dataset covers part of the National University of Singapore (NUS), Singapore. Hereafter, it is referred to as NUS dataset. The NUS dataset (see Fig. 3) was acquired by a terrestrial laser scanner (TLS) during the leaf-on season. The NUS campus has a very large coverage of greeneries, leading to the prevalence of occlusions during the process of collecting point clouds. Buildings in NUS dataset are generally lower than 30 m. However, the NUS site sits on a terrain with a relief of more than 60 m, making it possible to place TLS on high grounds to capture a comprehensive view of the site, including roofs of buildings. The dataset has 14 452 784 points and comes with color information. The study discerns the following five classes in NUS dataset: roof; wall; grass; tree; and ground.

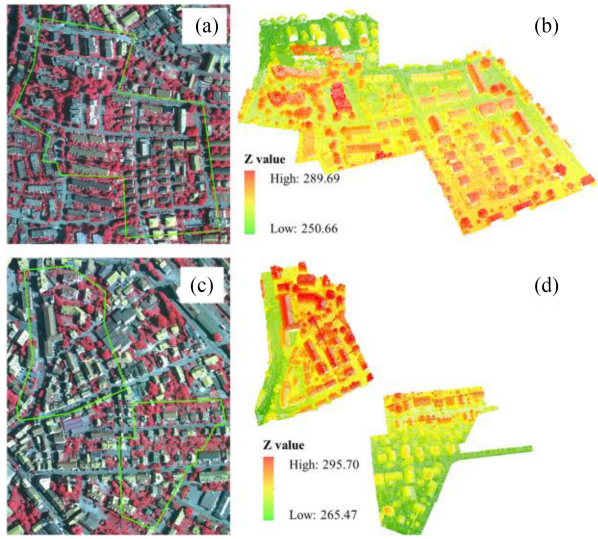

Fig. 4. ISPRS benchmark dataset of Vaihingen, Germany. (a) Orthoimage of area 1. (b) LiDAR point cloud of area 1. (c) Orthoimage of area 2. (d) LiDAR point cloud of area 2.

The second dataset covers Vaihingen, Germany, which is provided by International Society of Photogrammetry and Remote Sensing (ISPRS) benchmark. Hereafter, it is referred to as Vaihingen dataset. This dataset contains two sites (areas 1 and 2), which represent two typical scenes in Vaihingen. Area 1 is a residential community with small detached houses and a few high-rise buildings, which are surrounded by trees. Area 2 is situated in the center of the city and consists of dense buildings with complex shapes. Fig. 4 shows the neighborhoods of areas 1 and 2, and the LiDAR point clouds used in this article. The LiDAR points were acquired by an airborne laser scanner in August 2008. All points in the two areas have been manually labeled as one of the following six categories: grassland; impervious surface (IS); roof; facade; shrub/tree; and clutter. It has 1165598 points.

B. Results of 3-D Spatial Contextual Features

After point cloud segmentation and initial classification, the 3-D contextual features of each point segment were extracted. These features were analyzed by averaging feature values of training samples of each class. Figs. 5 and 6 show the 3-D contextual features by semantic classes of NUS dataset and Vaihingen dataset, respectively. Each figure has five subfigures illustrating

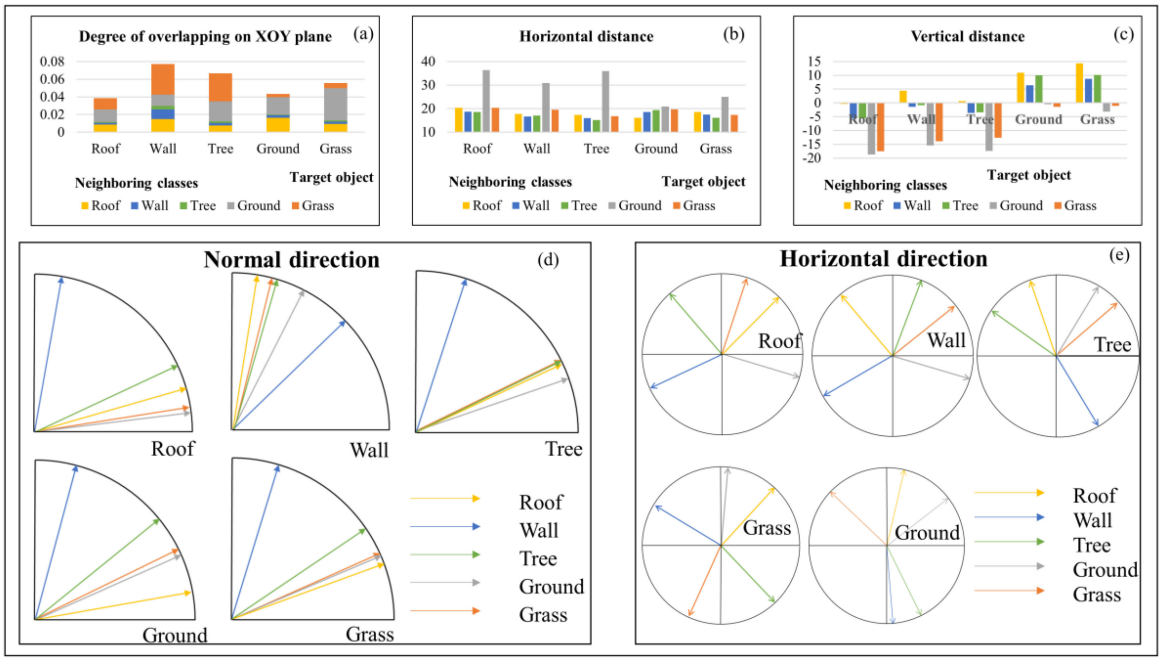

Fig. 5. 3-D contextual features of NUS dataset.

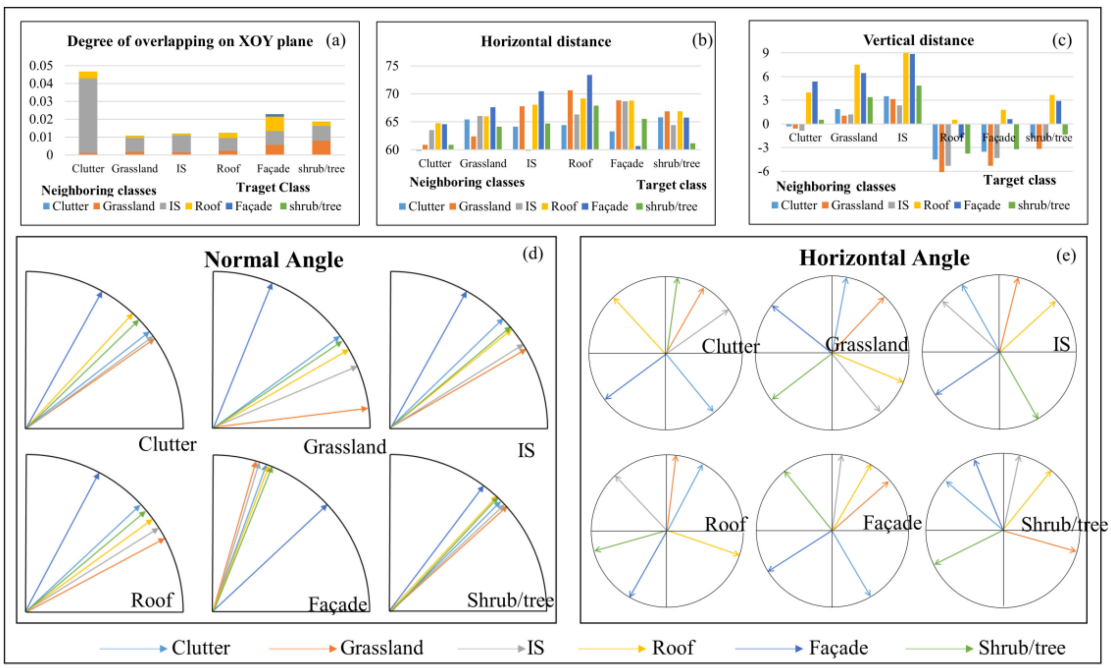

Fig. 6. 3-D contextual features of Vaihingen dataset.

1) DoV between a target segment and its neighboring segments.
2) Horizontal distances from a target segment to its neighboring segments.
3) vertical distances from a target segment to its neighboring segments.
4) Normal directions (i.e., the angle between the normal vectors of a target segment and of certain neighboring segments).
5) Horizontal directions (i.e., the angle of the horizontal direction of a target segment to neighboring segments.

The range of the vertical direction is [0, 90], while the range of the horizontal direction is [0, 360).

1) 3-D Contextual Features of NUS Dataset: For NUS dataset, a total of 25 (5 relationships $\times$ 5 classes) 3-D contextual features were extracted. The analysis results suggest the followings (see Fig. 5). First, a wall segment usually has high degrees of overlap with its neighboring segments, especially

grass segments. The same is found for tree segments. A segment of roof, ground, or grass also has high degrees of overlap with its neighboring segments but with ground segments only [see Fig. 5(a)]. Second, the segments of all classes except tree are distant, horizontally, from their neighboring ground segments [see Fig. 5(b)], which is consistent with the observation of rich vegetation within the NUS campus. Third, a roof segment is usually the highest among its neighboring segments, and has large vertical distances with its neighboring ground and grass segments and small vertical distances with its neighboring wall and tree segments [see Fig. 5(c)]. Fourth, a roof segment has an approximate 90° of normal direction to its neighboring wall segments, indicating that the neighboring wall objects are nearly perpendicular to the roof object. Similarly, for a wall segment, it is usually perpendicular to its neighboring roof and grass segments. For a ground segment, it is also usually perpendicular to its neighboring wall objects [see Fig. 5(d)]. Last, there exist alignment patterns for different classes in NUS dataset, which can be mined using the horizontal direction features [see Fig. 5(e)].

2) 3-D Contextual Features for Vaihingen Dataset: For Vaihingen dataset, a total of 30 (5 relationships × 6 classes) 3-D contextual features were extracted for each segment. Although the classification and the environment of Vaihingen dataset are different from that of NUS dataset, there are common characteristics. First, segments of all classes except facade have large degrees of overlap with their neighboring IS segments. Facade segments have large degrees of overlap with neighboring roof segments [see Fig. 6(a)]. Second, a clutter segment usually has a short horizontal distance with its neighboring segments while a segment of roof usually has a long horizontal distance with its neighboring segments. A short horizontal distance is also observed between segments of clutter and shrub/tree, indicating a strong spatial correlation between the two classes [see Fig. 6(b)]. Third, segments of roof, facade, and shrub/tree are usually higher than the neighboring segments of the remaining classes (i.e., clutter, grassland and IS) [see Fig. 6(c)]. Fourth, facade segments are nearly perpendicular to the neighboring segments of all classes except clutter [see Fig. 6(d)]. Fifth, the horizontal direction information in Vaihingen dataset is similar to that of the NUS dataset, which will be further mined and employed in the process of refined classification [see Fig. 6(e)].

The analysis of 3-D contextual features confirmed the usefulness of these features in distinguishing different classes. We also find that much of the obtained information from the 3-D contextual features concurs with commonsense knowledge. Although existing studies have attempted to adopt such heuristics for classifying point clouds (e.g., [15]), the presented approach extracts the information automatically and utilizes it quantitatively. The 3-D contextual information mined by the proposed method is also more comprehensive and robust.

C. Classification Results

A total of 11 584 and 14 024 segments were generated for the NUS and Vaihingen dataset. The two datasets were trained and classified separately as they have different category systems and point characteristics. To train each dataset, 3475 segments and 4207 segments were chosen from NUS dataset and Vaihingen dataset individually. The two datasets were evaluated at different levels. For NUS dataset, the evaluation was conducted at segment level because it is large and we do not have standard labels for each point. Instead of checking individual points, the classification result was assessed quantitatively at segment level by randomly selecting about 1200 segments and comparing their predicted labels with referenced labels identified manually. For Vaihingen dataset, as the standard label of each point is known, point-level evaluation, which is based on the comparison of the predicted label of each point in the test data (more than 800 000 points) with its reference label, was conducted. The refined results of 3-D point cloud classification of NUS dataset and Vaihingen dataset are shown in Figs. 7 –9, respectively. Both completeness and correctness for each class were evaluated, and F_1-score (F_1) was computed as a combination of the above two measures.

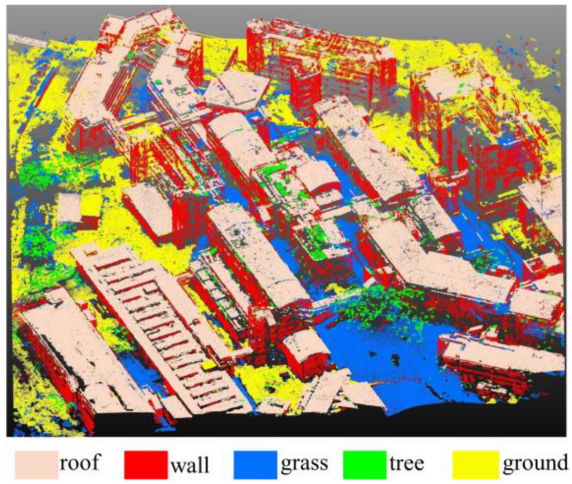

Fig. 7. Refined classification results of NUS dataset by the proposed method.

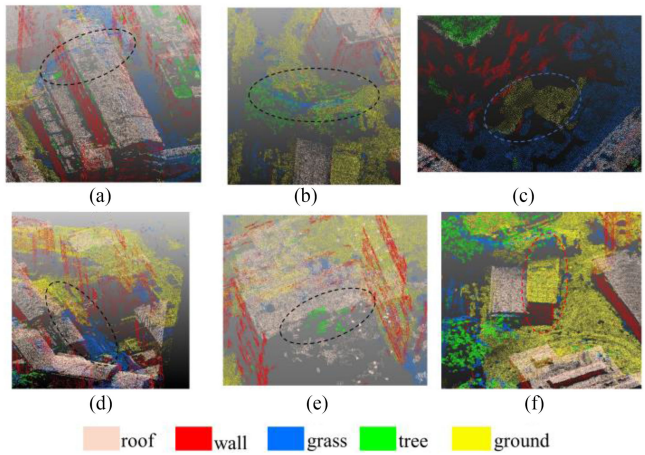

Fig. 8. Detailed classification results of certain areas in NUS dataset. (a) and (b) Misclassifications between grass and tree. (c) and (d) Misclassifications between grass and ground. (e) and (f) Misclassifications between tree and roof.

1) Classification Results of NUS Dataset: As can be seen in Fig.7, the refined classification result of NUS dataset is

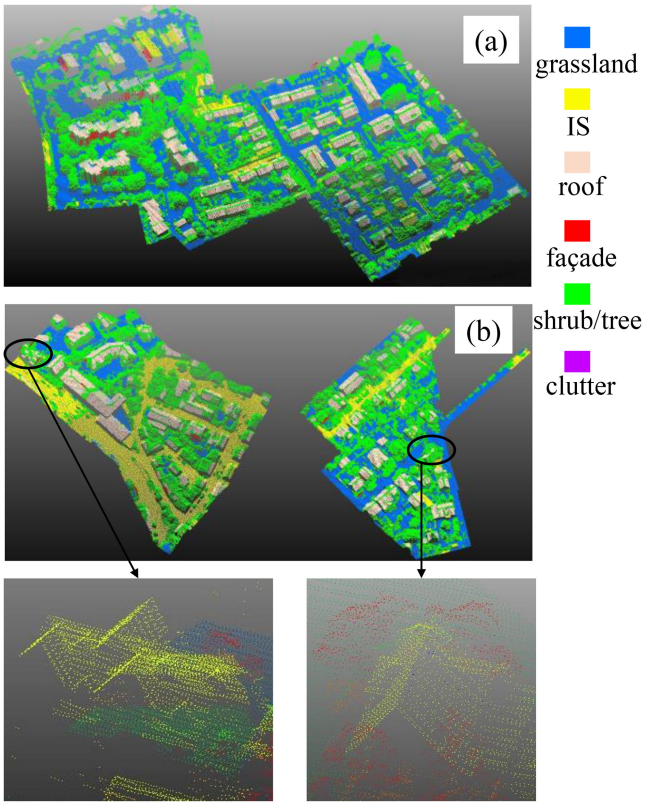

Fig. 9. Classification results of Vaihingen dataset by H-MLP.

TABLE II
QUANTITATIVE CLASSIFICATION RESULTS OF NUS DATASET AT SEGMENT LEVEL

	Refined classification by the proposed method (%)			Initial classification (%)		
Class	Cor	Com	F_1	Cor	Com	F_1
Roof	93.92	95.21	94.56	88.51	87.92	88.22
Wall	95.48	97.05	96.26	93.23	95.38	94.29
Grass	83.69	81.94	82.81	75.89	72.79	74.31
Tree	91.40	91.64	91.52	85.16	85.16	85.16
Ground	95.76	92.94	94.33	90.91	90.91	90.91
	OA = 92.51%; kappa index =0.90			OA = 87.46%; kappa index =0.84		

satisfactory. The quantitative evaluation, given in Table II, shows that the proposed method achieves an OA of 92.51% and a Kappa index of 0.90. The correctness, completeness, and F_1 values for each class show that classes of man-made objects such as roof, wall, and ground have higher classification accuracies, while classes of natural objects such as grass and tree have lower classification accuracies. The class wall has the highest completeness (97.05%) and F_1(96.26%) and the second highest correctness (95.48%). The high classification accuracy can be explained easily as most wall segments are vertical and thus they are unlikely misclassified as other categories. The class roof has the second highest F_1 (94.56%). Generally, roof segments have several distinctive geometrical features, specifically area and height, and most roof segments are flat in NUS dataset. However, with only these primitive features, misclassification between roof and other categories is frequent [15]. By incorporating spatial contextual features, many misclassifications are avoided. With the accurately classified wall segments, and the perpendicular relationships between neighboring wall and roof segments, the accuracy of classifying roof segments can be significantly improved. The lowest and second lowest F_1 were obtained by the grass (82.81%) and the tree class (91.52%).

According to Fig. 7 and the detailed classification results shown in Fig.8, we found some misclassifications. The first type of misclassification occurred between the grass and the tree class [see Fig. 8(a) and 8(b)], which was because these two classes have very similar spectral features. Aside from the above confusions, some grass segments were wrongly labeled as ground [see Fig. 8(c) and 8(d)], and some roof segments were wrongly labeled as tree [see Fig. 8(e)] or ground [see Fig. 8(f)]. We found that many of the above confusions were caused by some similar geometrical features these classes share. For example, both grass and ground usually have low heights, and roof and tree segments generally have larger heights. Another reason for the misclassification between the tree and roof class might be the occlusions of roof because of the nearby trees, which would result in incomplete geometries of roof segments.

2) Effectiveness of 3-D Contextual Features in NUS Dataset Classification: In this section, the refined classification results with contextual features and the initial classification results without contextual features are compared. The initial classification by random forest achieved an OA of 87.46% and a kappa index of 0.84 (see Table II). The results indicate that the additional spatial contextual features improve overall classification accuracy by around 5%. The improvement for individual classes measured by F_1 is between 1.97% and 8.5%, where the class grass has the highest improvement percentage while the class wall has the least improvement. The result suggests that incorporating spatial contextual features benefits more for classifying natural than man-made objects, which differs somewhat from those by Niemeyer *et al*. [20] where the highest improvement was achieved for the class facade.

3) Classification Results of Vaihingen Dataset: Fig. 9(a) and (b) shows classification results in Area 1 and Area 2 of Vaihingen dataset. Fig. 10 shows discrepancies between H-MLP and initial classification, in which red represents points that were recognized correctly by H-MLP but incorrectly in the results of initial classification that did not consider spatial contextual features. As can be seen in Figs. 9 and 10, most of the points are correctly classified. Quantitatively, the proposed method achieved an OA of 82.34% and a kappa index of 0.74 (see Table III). The lower OA of Vaihingen dataset than that of NUS dataset might be due to evaluation levels (point- versus segment-level). It is worth mentioning that misclassification due to the imperfect segmentation results is unavoidable, thus causing the segment-level evaluation scores generally higher than those of point-level evaluation. The correctness, completeness, and F_1 of each class is also given in Table III. The highest F-score was obtained for the class roof (94.75%), with both correctness and completeness above 90%. The lowest F_1 was obtained for the class clutter (55.18%), which can be largely attributed to its low completeness of only 40.11%. As the class clutter represents a combination of powerline, car,

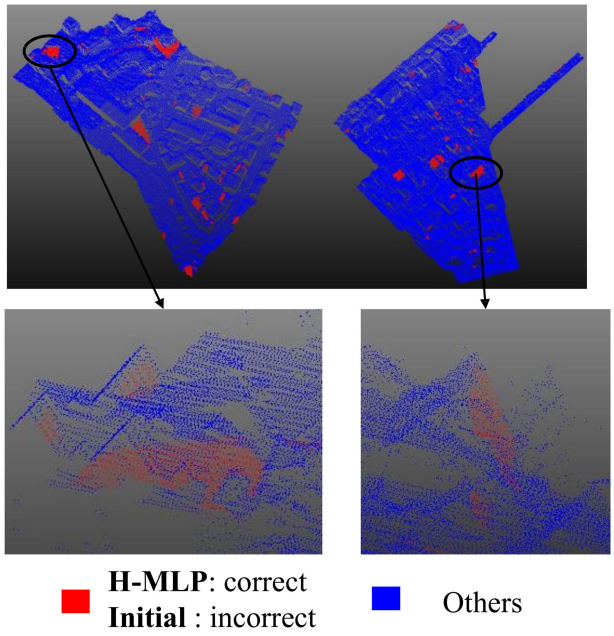

Fig. 10. Discrepancies between H-MLP and initial classification for Vaihingen dataset.

TABLE III
QUANTITATIVE CLASSIFICATION RESULTS OF VAIHINGEN DATASET AT POINT LEVEL

	Refined classification by our method (%)			Initial classification (%)		
Class	Cor	Com	F_1	Cor	Com	F_1
Clutter	88.40	40.11	55.18	37.99	35.35	36.62
Grassland	89.84	55.01	68.24	89.69	53.78	67.24
IS	71.99	98.70	83.25	68.67	99.12	81.13
Roof	97.42	92.22	94.75	96.29	82.47	88.85
Facade	70.82	66.64	70.82	77.49	64.46	70.38
Shrub/tree	80.49	48.17	60.27	74.24	47.33	57.81
	OA = 82.34%; kappa index =0.74			OA = 79.22%; kappa index =0.69		

and fence/hedge, its intraheterogeneity is large, and thereby low evaluation scores. Another example is the relatively low F_1 of class facade (70.82%) compared to the F_1 of the same class in NUS dataset (96.26%), which reflects the sparse points issue of airborne LiDAR and thus incomplete structures for facades. Nonetheless, the detection of facades can be considered good compared to its much lower evaluation scores achieved by other published methods (see Table VI). In addition, significant confusions between grassland and IS are observed. We consider spectral features are crucial in differentiating these two classes. However, the only spectral features in Vaihingen dataset are mean intensity and standard intensity scores.

4) Effectiveness of 3-D Contextual Features in Vaihingen Dataset Classification: Similar to NUS dataset, the refined classification results with contextual features and the initial classification results without contextual features are compared. The quantitative evaluations show that the initial classification achieved an OA of 79.22% and a kappa index of 0.69 (see Table III). The proposed method improved OA by 3.12% and

TABLE IV
OVERVIEW OF THE 15 MOST IMPORTANT FEATURES BASED ON GAIN RATIO VALUES

Rank	Importance (%)	Feature	Type
1	5.45	Height above grassland	Contextual (distance)
2	4.32	Normal value in y direction	Primitive
3	4.25	Including angle with roof	Contextual (direction)
4	3.87	Overlapping degree with facade on XOY plane	Contextual (topology)
5	3.85	Height above shrub/tree	Contextual (distance)
6	3.42	Height above roof	Contextual (distance)
7	3.40	Normal value in z direction	Primitive
8	3.21	The third eigenvalue	Primitive
9	3.09	Vertical direction with grassland	Contextual (direction)
10	3.08	Height above impervious surface	Contextual (distance)
11	3.01	Anisotropy	Primitive
12	3.01	Sphericity	Primitive
13	2.97	Mean intensity value	Primitive
14	2.95	Surface variation	Primitive
15	2.89	Eigentropy	Primitive

TABLE V
OVERVIEW OF THE 20 MOST IMPORTANT BUT UNCORRELATED FEATURES BASED ON CFS

Rank	Feature	Type
1	Height above grassland	Contextual (distance)
2	Normal value in y direction	Primitive
3	Including angle with roof	Contextual (direction)
4	Height above shrub/tree	Contextual (distance)
5	Height above roof	Contextual (distance)
6	Normal value in z direction	Primitive
7	Height above impervious surface	Contextual (distance)
8	Anisotropy	Primitive
9	Mean intensity value	Primitive
10	Surface variation	Primitive
11	Vertical direction with shrub/tree	Contextual (direction)
12	Vertical direction with impervious surface	Contextual (direction)
13	Number of points in a segment	Primitive
14	Height above facade	Contextual (distance)
15	Omnivariance	Primitive
16	Standard deviation of intensity value	Primitive
17	Overlapping degree with roof on XOY plane	Contextual (topology)
18	Height above clutter	Contextual (distance)
19	Overlapping degree with grassland on XOY plane	Contextual (topology)
20	Horizontal distance with roof	Contextual (distance)

TABLE VI
QUANTIFICATION COMPARISON OF THE PROPOSED METHOD (H-MLP) AND MLP AND OTHER PUBLISHED APPROACHES ON THE ISPRS BENCHMARK DATASET

Method	Clutter			grassland	IS	Roof	Façade	Shrub/tree		OA
	Powerline	Car	Fence/Hedge	grassland	IS	Roof	Façade	Shrub	Tree	
H-MLP		55.2		68.2	83.3	**94.7**	**70.8**	60.3		**82.3**
MLP		45.3		65.8	81.9	92.6	**70.9**	59.8		80.3
S3*	54.4	57.9	28.9	65.2	85.0	90.9	-	39.5	75.6	76.2
MSMTN*	5.9	30.1	16.0	20.1	61.0	60.7	42.8	32.5	64.2	50.2
HHO-CRF*	-	**73.1**	34.0	77.5	**91.1**	94.2	56.3	46.6	**83.1**	81.6
CNN*	37.1	63.4	23.9	**81.4**	90.1	93.4	47.5	39.9	78.0	**82.3**
MCNN*	-	**66.7**	40.7	**88.8**	**91.2**	**93.6**	42.6	55.9	**82.5**	**85.2**

F_1 of Each Class Together With OA are Presented in percentage. The Highest and Second Highest Scores are in Bold. (*Results Acquired From http://www2.isprs.org/commissions/comm2/wg4/vaihingen-3d-semantic-labeling.html).

the kappa index by 0.05. The improvement does not seem to be significant at the first glance. However, the evaluation scores for some classes are greatly improved. The highest F_1 improvement of 18.56% was obtained by the class clutter, while the least F_1 improvement of 0.44% was obtained by the class facade. Aside from the class clutter, evaluation scores of the class roof are greatly improved with an increase of 5.90% in F_1. Although most correctness and completeness values obtained from the results of the proposed approach improved from the initial classification, the correctness of facade dropped from 77.49% to 70.82%. At the same time, the completeness of facade is increased from 64.66% to 66.64%. Overall, the proposed method and the additional 3-D spatial contextual features have indeed improved the results of the 3-D point cloud classification for typical urban classes.

5) Computational Performance: A prototype based on C++ and Python was developed for the proposed method. All of the experiments were performed on a desktop computer with a CPU Intel Core i7-6700 processor, with 3.4 GHz and 32-GB RAM. For NUS dataset with 14452784 points, the elapsed time for point segmentation, initial classification and H-MLP training is nearly 16.3 min, 3.7 min, and 4.8 h.

IV. Discussions

A. Importance Evaluation of Contextual Features

To further verify the effectiveness of additional 3-D spatial contextual features, the importance of these features is assessed using gain ratio [39] as it is commonly used for assessing the importance of a single feature, and has been successfully employed in previous studies [15], [40]. Here, we evaluate the feature importance on Vaihingen dataset.

A total of 50 features were extracted in the Vaihingen dataset, in which 30 were 3-D spatial contextual features and 20 were primitive features. Table IV gives the top 15 most important features based on their gain ratio values in percentage. Among these features, height above grassland is the most important, contributing 5.45% to the classification. Almost half of these features (7 out of 15) are 3-D spatial contextual features. In addition, the total contribution of the seven 3-D spatial contextual features (27.0%) is larger than that of the eight primitive features (25.76%). The large ratio and high contribution of 3-D spatial contextual features demonstrate their importance in the classification of point clouds. Aside from height above grassland, height above roof, shrub/tree, roof and IS are included in the 15 most important features, indicating the importance of vertical distance features in distinguishing different classes. However, the feature of absolute height does not seem to contribute a lot (rank 34 with an importance of 1.4%). The reason might be that the study area of the Vaihingen dataset sits on an uneven terrain where height is not able to function well in distinguishing different objects. Among the seven contextual features in Table IV, the numbers of topological, distance and directional features are 1, 4, and 2, and their accumulated contributions are 3.87%, 15.8% and 7.34%.

In addition to the assessment of single feature contribution, the correlation between features is another important factor for evaluating feature importance. In this article, we used correlation-based feature selection (CFS) [40] to select the most important but uncorrelated features from the full feature set. A total of 20 most important features were selected (see Table V). Most of these features overlap with the features in Table IV, except for overlapping degree with façade on XOY plane, the third eigenvalue, vertical direction with grassland, sphericity and eigentropy. This is because these features have high correlations with other more important features. Of all the 20 features selected by CFS, 12 of them are contextual features, in which the numbers of topological, distance and directional features are 2, 7, and 3. The correlation analysis verified that the 3-D spatial contextual features can indeed bring additional information and contribution for discerning different classes in point cloud classification.

B. Comparisons Between H-MLP and MLP

The quantitative classification results in Section III-C and feature importance evaluation results in Section IV-A have both indicated the great significance of incorporating additional 3-D spatial contextual features. To validate the performance of H-MLP, we compared the classification results of Vaihingen dataset obtained by separating contextual features and primitive features to different levels using the proposed H-MLP network to those by treating all features equally in the MLP network.

Table VI gives the quantitative results by H-MLP and MLP. Concerning OA, the H-MLP (82.3%) outperformed MLP (80.3%) by 2.0%. In addition, the proposed method achieved a higher F_1 than MLP for almost all classes, except for facade whose F_1 is lower than that by MLP by merely 0.1%. The results demonstrate that the hierarchical structure of primitive and 3D contextual features helps improve the classification accuracy.

C. Comparisons With Other Published Methods using Vaihingen Dataset

For further evaluation, we compared our results with the results of existing methods including supervoxel-based spectro-spatial approach (S3) by Ramiya *et al*. [41]; multiscale, multitype neighborhood (MSMTN) by Blomley *et al*. [42]; hierarchical higher order CRF (HHO-CRF) [43]; convolutional neural network (CNN) by Yang *et al.* [44]; and multiscale convolutional neural network (MCNN) by Zhao *et al*. [45]. S3 first segments 3-D point cloud into supervoxels, which are further classified by different machine learning techniques with spectral and geometrical features. MSMTN extracts complementary types of geometric features from multiscale and multitype neighborhoods. The extracted features are set as input to several classifiers with different learning principles. HHO-CRF is an improved version of the work in Niemeyer *et al.* [20], which integrates a Random Forest classifier into a CRF framework. The HHO-CRF incorporates spatial and semantic context via a two-layer CRF: the first layer operates on a point level and the second layer operates on a segment level. The MCNN is a CNN-based method that extracts high-level representation of features and labels each 3-D point individually. MCNN first creates a group of multiscale contextual images for each point. Second, these

contextual images are input to a multi-scale neural network to learn deep features. Finally, the data with a combination of deep features from multiple scales is trained and classified using a softmax regression classifier.

The results by the above existing methods are shown in Table VI. The category system of our study is slightly different from that of others in class clutter and shrub/tree (see Table VI), in which we discern a total of six classes, while they discern nine classes. The OA of the proposed method ranks second (82.3%), and is 2.9% lower than the highest OA achieved by MCNN. Additionally, our method achieves the highest F_1 in terms of class roof and facade. The F_1 of class facade by our method (70.8%) is 14.5% higher than the second best (56.3%) produced by HHO-CRF. The F_1 of class roof by our method (94.7%) has a slight increase of 1.1% from the second best (93.6%), again achieved by MCNN. It is also worth noting that F_1 of other classes by our method are lower than those by other methods, especially for class grassland and IS, which are 20.6% and 7.9% lower than the best one. By visual inspection, we find that the relatively low F_1 of the two classes are largely attributed to the mis-segmentation results (see Fig. 9). The individual F_1 and OA can expect higher scores by improving the segmentation results in the future.

V. Conclusion

In this article, a novel approach, which adopts a hierarchical classifier and 3-D contextual features to improve the classification accuracy of 3-D point cloud was proposed. It comprises three main steps. First, initial segment-based classification was conducted by extracting primitive features of individual segments and adopting a conventional RF classifier. Second, based on the initial classification results, novel 3-D contextual features for each point segment were extracted, which consider multiple types of spatial relationships between neighboring segments and semantic dependencies. Third, an H-MLP classifier, which considers primitive features and spatial contextual features at different levels, was proposed to reclassify the point cloud data and refine the initial classification results.

Two experiments were performed with a terrestrial laser scanning based NUS dataset and an aerial laser scanning based Vaihingen dataset provided by ISPRS. The quantitative evaluation showed that the additional 3-D spatial contextual features improve the classification accuracy significantly. The OA for NUS dataset increased by 5.05% when discerning five typical classes, while an improvement of 3.12% was observed on the more complex Vaihingen dataset discerning six classes. Additionally, the contribution of 3-D contextual features to classification was evaluated. By evaluating the importance of single features, the results showed that the total contribution of contextual features almost equaled those of the primitive features. The importance of the feature set was also evaluated, which considers the correlation between features. The results showed that the correlation between spatial contextual features and primitive features was minimal, and the contextual features indeed brought additional information useful for discriminating different classes of point clouds. Quantitative classification results of Vaihingen dataset by the proposed H-MLP were compared with the conventional MLP network and other existing methods. The OA increased by 2.0% from the conventional MLP classifier, indicating the advantage of the proposed hierarchical network. The comparisons with other existing methods showed that the OA by the proposed method ranked second (82.3%), which was 2.8% lower than the state-of-art performance. However, the F_1 of class roof (94.7%) and façade (70.8%) achieved the best results.

While the quantitative evaluation demonstrated the effectiveness of the proposed method, there are still three limitations. First, the classification results are highly dependent on the segmentation results, and undersegmentation or oversegmentation are unavoidable. Second, the accuracy of the 3-D spatial contextual features highly relies on the initial classification results. Third, some contextual features are ineffective for distinguishing different categories, and it may help improve the efficiency and accuracy of the final classification by automatically removing those contextual features. In the future, we will extend our research to overcome these limitations.

References

[1] Q. Zhou and U. Neumann, "Complete residential urban area reconstruction from dense aerial lidar point clouds," *Graphical Models*, vol. 75, pp. 118–125, May 2013.

[2] A. Sampath and J. Shan, "Urban DEM generation from raw LiDAR data: A labeling algorithm and its performance," *Photogramm. Eng. Remote Sens.*, vol. 71, pp. 217–226, Feb. 2005.

[3] D. Horvat, B. Zalik, and D. Mongus, "Context-dependent detection of non-linearly distributed points for vegetation classification in airborne LiDAR," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 1–14, Jun. 2016.

[4] E. Kwak and A. Habib, "Automatic representation and reconstruction of DBM from LiDAR data using recursive minimum bounding rectangle," *ISPRS J. Photogramm. Remote Sens.*, vol. 93, pp. 171–191, Jul. 2014.

[5] A. Boyko and T. Funkhouser, "Extracting roads from dense point clouds in large scale urban environments," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, pp. 502–512, Dec. 2011.

[6] J. Han, D. Kim, M. Lee, and M. Sunwoo, "Enhanced road boundary and obstacle detection using a downward-looking LiDAR sensor," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 971–985, Mar. 2012.

[7] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 286–304, Jul. 2015.

[8] Q. Zhu, Y. Li, H. Hu, and B. Wu, "Robust point cloud classification based on multi-level semantic relationships for urban scenes," *ISPRS J. Photogramm. Remote Sens.*, vol. 129, pp. 86–102, Jul. 2017.

[9] J. Eum, M. Bae, J. Jeon, H. Lee, S. Oh, and M. Lee, "Vehicle detection from airborne LiDAR point clouds based on a decision tree algorithm with horizontal and vertical features," *Remote Sens. Lett.*, vol. 8, pp. 409–418, May 2017.

[10] J. Lopatin, K. Dolos, H. J. Hernández, M. Galleguillos, and F. E. Fassnacht, "Comparing generalized linear models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile," *Remote Sens. Environ.*, vol. 173, pp. 200–210, Feb. 2016.

[11] C. Chen *et al.*, "The mixed kernel function SVM-Based point cloud classification," *Int. J. Precis. Eng. Manuf.*, vol. 20, pp. 737–747, Mar. 2019.

[12] A. Maligo and S. Lacroix, "Classification of outdoor 3D lidar data based on unsupervised Gaussian mixture models," *IEEE Trans. Automat. Sci. Eng.*, vol. 14, no. 1, pp. 5–16, Jan. 2016.

[13] W. Zhang, Y. Guo, M. Lu, and J. Zhang, "Ground target detection in lidar point clouds using adaboost," in *Proc. Int. Conf. Control, Automat. Inf. Sci.*, 2015, pp. 1–7.

[14] K. Khoshelham and S. J. Oude Elberink, "Role of dimensionality reduction in segment-based classification of damaged building roofs in airborne laser scanning data," in *Proc Int. Conf. Geographic Object Based Image Anal.*, 2012, pp. 372–277.

[15] C.-C. Feng and Z. Guo, "Automating parameter learning for classifying terrestrial LiDAR point cloud using 2D land use maps," *Remote Sens.*, vol. 10, pp. 1192–1214, 2018.
[16] L. Gao, T. Li, L. Yao, and F. Wen, "Research and application of data mining feature selection based on relief algorithm," *J. Softw.*, vol. 9, pp. 515–522, Feb. 2014.
[17] F. Koutanaei, H. Sajedi, and M. Khanbabaei, "A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring," *J. Retailing Consum. Serv.*, vol. 27, pp. 11–23, Nov. 2015.
[18] M. A. Hall and G. Holmes, "Benchmarking attribute selection techniques for discrete class data mining," *IEEE Trans. Geosci. Remote Sens.*, vol. 15, no. 6, pp. 1437–1447, Nov./Dec. 2003.
[19] M. Najafi, S. Namin, M. Salzmann, and L. Petersson, "Non-associative higher-order Markov networks for point cloud classification," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 500–515.
[20] 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.
[21] D. Munoz, J. Bagnell, N. Vandapel, and M. Hebert, "Contextual classification with functional max-margin Markov networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 975–982.
[22] R. Shapovalov, A. Velizhev, and O. Barinova, "Non-associative markov networks for 3D point cloud classification," in *Proc. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, 2010, pp. 103–108.
[23] S. Kumar and M. Hebert, "Discriminative random fields," *Int. J. Comput. Vis.*, vol. 68, pp. 179–201, Jun. 2006.
[24] S. Laible, Y. Khan, and A. Zell, "Terrain classification with conditional random fields on fused 3D LiDAR and camera data," in *Proc. Eur. Conf. Mobile Robots,* 2013, pp. 172–177, .
[25] J. Niemeyer, J. Wegner, C. Mallet, F. Rottensteiner, and U. Soergel, "Conditional random fields for urban scene classification with full waveform LiDAR data," in *Proc. ISPRS Conf. Photogramm. Image Anal.*, 2011, pp. 233–244.
[26] J. Niemeyer, F. Rottensteiner, and U. Soergel, "Conditional random fields for LiDAR point cloud classification in complex urban areas," in *Proc. ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, 2012, pp. 263–268.
[27] Z. Li *et al.*, "A three-step approach for TLS point cloud classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5412–5424, Sep. 2016.
[28] L. Landrieu, H. Raguet, B. Vallet, C. Mallet, and M. Weinmann, "A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 102–118, Oct. 2017.
[29] R. Huang, Y. Xu, D. Hong, W. Yao, P. Ghamisi, and U. Stilla, "Deep point embedding for urban classification using ALS point clouds: A new perspective from local to global," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 62–81, May 2020.
[30] N. Li, C. Liu, and N. Pfeifer, "Improving LiDAR classification accuracy by contextual label smoothing in post-processing," *ISPRS J. Photogramm. Remote Sens.*, vol. 148, pp. 13–31, Feb. 2019.
[31] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller, "Multi-class segmentation with relative location prior," *Int. J. Comput. Vis.*, vol. 80, no. 3, pp. 300–316, Apr. 2008.
[32] 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.*, 2009, pp. 2154–2161.
[33] B. Yang, Z. Dong, Y. Liu, F. Liang, and Y. Wang, "Computing multiple aggregation levels and contextual features for road facilities recognition using mobile laser scanning data," *ISPRS J. Photogramm. Remote Sens.*, vol. 126, pp. 180–194, Apr. 2017.
[34] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4490–4499.
[35] C. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," *IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, no. 2, pp. 652–660, Apr. 2017.
[36] C. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical features learning on point sets in a metric space," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5099–5108.
[37] 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.
[38] J. Papon, A. Abramov, M. Schoeler, and F. Worgotter, "Voxel cloud connectivity segmentation-supervoxels for point clouds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013. pp. 2027–2034.
[39] J. Quinlan, "Improved use of continuous attributes in C4.5," *J. Artif. Intell. Res.*, vol. 4, pp. 77–90, Mar. 1996.
[40] L. Ma, L. Cheng, M. Li, and X. Ma, "Training set size, scale, and features in geographic object-based image analysis of very high resolution unmanned aerial vehicle imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 102, pp. 14–27, Apr. 2015.
[41] A. Ramiya, R. R. Nidamanuri, and K. Ramakrishnan, "A supervoxel-based spectro-spatial approach for 3D urban point cloud labelling," *Int. J. Remote Sens.*, vol. 37, pp. 4172–4200, Jul. 2016.
[42] R. Blomley, B. Jutzi, and M. Weinmann, "Classification of airborne laser scanning data using geometric multi-scale features and different neighbourhood types," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 3, no. 3, Jun. 2016.
[43] J. Niemeyer, F. Rottensteiner, U. Soergel, and C. Heipke, "Hierarchical higher order CRF for the classification of airborne LiDAR point clouds in urban areas," in *Proc. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, 2016, pp. 655–662.
[44] Z. Yang, W. Jiang, B. Xu, Q. Zhu, S. Jiang, and W. Huang, "A convolutional neural network-based 3D semantic labeling method for ALS point clouds," *Remote Sens.*, vol. 9, no. 9, pp. 936, Sep. 2017.
[45] R. Zhao, M. Pang, and J. Wang, "Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network," *Int. J. Geographical Inf. Sci.*, vol. 32, pp. 960–975, Feb. 2018.
[46] S. Aksoy, C. Tusk, K. Koperski, and G. Marchisio, "Scene modeling and imaging mining with a visual grammar," in *Frontiers of Remote Sensing Information Processing*, C. Chen, Ed., Singapore: World Sci., 2003, pp. 35–62.
[47] Q. Cheng, J. Wang, J. Shang, and B. Daneshfar, "Spatial relationship-assisted classification from high-resolution remote sensing imagery," *Int. J. Digit. Earth*, vol. 8, no. 9, pp. 710–726, Jan. 2015.
[48] M. Pauly, M. Gross, and L. P. Kobbelt, "Efficient simplification of point sample surface," in *Proc. Conf. Vis.*, 2002, pp. 163–170.

Chen-Chieh Feng received the B.S. and M.S. degrees in geography from National Taiwan University, Taipei, Taiwan, and the Ph.D. degree in geography from the State University of New York, University at Buffalo, Buffalo, NY , USA, in 1994, 1996, and 2004, respectively.

He is currently an Associate Professor with the Department of Geography, National University of Singapore, Singapore. His research interests include 3-D GIS and smart city, spatial data mining, and the use of GIS and remote sensing for studying land use land cover changes and evaluating ecosystem health.

Zhou Guo received the B.S. degree in remote sensing from Wuhan University, Wuhan, China, and the Ph.D. degree in GIS from Peking University, Beijing, China, in 2012 and 2017, respectively.

He was a Postdoctoral Fellow with the Department of Geography, National University of Singapore, Singapore, from 2017 to 2020. He is currently an Associate Professor with the School of Geospatial Engineering and Science, Sun Yat-Sen University, Guangzhou, China. His research interests include Lidar data processing, high-resolution satellite image classification, and smart city applications