This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Instance-Aware Semantic Segmentation of Road Furniture in Mobile Laser Scanning Data

Fashuai Li, Zhize Zhou, Jianhua Xiao, Ruizhi Chen, Matti Lehtomäki, Sander Oude Elberink, George Vosselman, Juha Hyyppä, Yuwei Chen, and Antero Kukko

***Abstract*— In this paper, we present an improved framework for the instance-aware semantic segmentation of road furniture in mobile laser scanning data. In our framework, we first detect road furniture from mobile laser scanning point clouds. Then we decompose the detected pieces of road furniture into poles and their attached components, and extract the instance information of the components with different features. Most importantly, we classify the components into different categories by combining a classifier and a probabilistic graphic model named DenseCRF, which is the major contribution of this paper. For the classification of the components using DenseCRF, the unary potentials and the pairwise potentials are first obtained. The unary potentials are obtained from the classifier which takes the instance information of components as the input. The pairwise potentials are calculated considering contextual relations between components. By utilising DenseCRF, the contextual consistency of components is preserved, and the performance is significantly improved compared to our previous work. We collect three datasets to test our framework, and compare the classification performances of six different classifiers with and without DenseCRF. The combination of random forest with DenseCRF outperforms the other methods and achieves high overall accuracies of 83.7%, 96.4% and 95.3% in these three datasets. Experimental results demonstrate that our framework reliably assigns both semantic information and instance information for mobile laser scanning point clouds of road furniture.**

***Index Terms*— Densely connected conditional random fields, instance-aware semantic segmentation, mobile laser scanning point clouds, pole-like road furniture.**

Manuscript received December 7, 2020; revised June 8, 2021, September 27, 2021, November 14, 2021, and February 27, 2022; accepted March 3, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2020AAA0108901; in part by the China Postdoctoral Science Foundation under Grant 2020M682505; in part by the Open Fund of State Key Laboratory of CAD&CG, Zhejiang University, under Grant A2022; in part by the Academy of Finland Flagship Project under Grant 337656; in part by the Strategic Research Council at the Academy of Finland Project under Grant 293389 and Grant 314312; and in part by the Academy Finland Project under Grant 300066. The Associate Editor for this article was D. F. Wolf. *(Corresponding author: Zhize Zhou.)*

Fashuai Li is with the State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058, China, also with the Wuhan Geomatics Institute, Wuhan 430022, China, and also with LIESMARS, Wuhan University, Wuhan 430022, China (e-mail: lifashuai@gmail.com).

Zhize Zhou is with the State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058, China (e-mail: zhouzhize@zju.edu.cn).

Jianhua Xiao is with the Wuhan Geomatics Institute, Wuhan 430022, China (e-mail: xjhwhkc@163.com).

Ruizhi Chen is with LIESMARS, Wuhan University, Wuhan 430022, China (e-mail: ruizhi.chen@whu.edu.cn).

Matti Lehtomäki, Juha Hyyppä, and Yuwei Chen are with the Finnish Geospatial Research Institute, Masala, 02430 Kirkkonummi, Finland, also with the Centre of Excellence in Laser Scanning Research, Academy of Finland, 00521 Helsinki, Finland, and also with the Department of Built Environment, Aalto University, 02150 Espoo, Finland (e-mail: matti.lehtomaki@nls.fi; juha.hyyppa@nls.fi; yuwei.chen@nls.fi).

Sander Oude Elberink and George Vosselman are with the Faculty of Geo-Information Science and Earth Observation, University of Twente, 7500 Enschede, The Netherlands (e-mail: s.j.oudeelberink@utwente.nl; george.vosselman@utwente.nl).

Antero Kukko is with the Finnish Geospatial Research Institute, Masala, 02430 Kirkkonummi, Finland, and also with the Centre of Excellence in Laser Scanning Research, Academy of Finland, 00521 Helsinki, Finland (e-mail: antero.kukko@nls.fi).

Digital Object Identifier 10.1109/TITS.2022.3157611

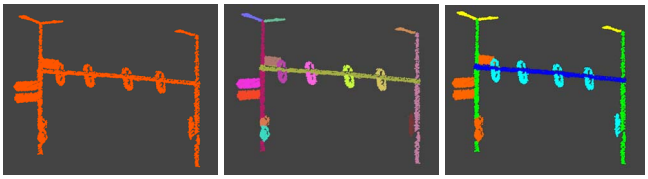
I. Introduction

ROAD furniture has long been playing an important role in traffic functionalities, and with the development of autonomous driving system, the automatic recognition of road furniture has become a prevalent research topic since the recognition precision will significantly influence the system reliability. To enhance the recognition, Stanford has made a priori list of traffic light locations so that its vehicle, Junior, can detect traffic lights in different lighting conditions. However, this kind of priori list relies heavily on manual interpretation, which is tedious and time-consuming. In this case, to automatically produce such a prior list of road furniture is in great demand.

Benefiting from the utilisation of mobile laser scanning (MLS) system which is broadly used to collect 3D point clouds of urban road scenes, significant progress has been made in the automatic recognition of road furniture in point clouds in recent years [1]–[3], [7], [10]. Specifically, since the rapid development of laser scanners has enabled MLS to capture denser and more accurate point clouds, road furniture recognition has therefore been advanced from the detection of pole-like road furniture to the recognition of different types of pole-like road furniture. However, the current granularity in pole-like road furniture recognition is still too coarse to meet the demand for precise and detailed mapping. Most studies simply classify pole-like road furniture as a whole, which is inappropriate for road furniture with mixed functionalities. For instance, a piece of road furniture with multiple classes can only be detected as a traffic light or a street light (Fig. 1a) by most previous work, whereas it should at least be decomposed as shown in Fig. 1b. Based on the functionalities, this piece of road furniture can be further interpreted as two vertical poles, one horizontal pole, three street lights, five traffic lights and five traffic-functionality signs (Fig. 1c), which is more appropriate for current demands of industrial applications such as 3D precise mapping.

1558-0016 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) (b) (c)

Fig. 1. A piece of road furniture with multiple functionalities. The original point cloud of one piece of road furniture (a) is decomposed into traffic-functionality components, where each colour denotes a component (b), and further interpreted based on their functionalities (c) (Orange: traffic-functionality signs, yellow: street lights, cyan: traffic lights, green: vertical poles, blue: horizontal poles).

In order to obtain a more fine-grained recognition, a piece of our early work about semantic segmentation of pole-like road furniture has been proposed in [2]. However, features were not well designed and contextual consistency between components was not exploited in our previous framework, which led to some misinterpretations of road furniture. In this paper, we thereby introduce densely connected conditional random field (DenseCRF), a probabilistic graphical model (PGM), and subsequently propose a novel framework based on it to improve the robustness of the previous recognition method. Besides, unary and pairwise potentials are adopted in DenseCRF to preserve both the mounting patterns of road furniture and the contextual consistency between components. In this framework, road furniture is first detected from MLS data. Then these pieces of road furniture are segmented into poles and their attached components at the level of instance information by a decomposition algorithm [4]. The instance information in this paper is to distinguish individual objects from each other – even if they belong to the same class. For example, traffic lights are not only assigned with class labels but each traffic light is also tagged with an identification number. As illustrated in Fig. 1, five traffic lights have been detected and segmented. The components are subsequently categorised based on their functionalities, using a combination of a simple classifier and DenseCRF, whereas the classifier can be any machine learning frameworks or knowledge-driven methods. In our experiment, five machine learning classifiers and a knowledge-driven classifier are tested with and without DenseCRF to verify the effectiveness of our framework, while deep neural network classifiers are not tested in this paper due to the limited number of poles' attached components. The results show that the performance of pole-like road furniture interpretation in MLS data has been significantly improved compared to our previous framework [2], and the combination of random forest (RF) classifier and DenseCRF performs the best among all our test cases.

The contributions of this paper are as follows:

- A novel framework is proposed for instance-aware sematic segmentation of pole-like road furniture, in which DenseCRF is introduced to improve the classification of poles' attached components.
- Unary potentials are adopted in DenseCRF to depict the mounting patterns of poles and their attached components, where new contextual features are designed as the input for their calculation. Besides, pairwise potentials are utilised in DenseCRF to preserve the contextual consistency between attached components.
- The state-of-the-art performance is achieved by our framework in terms of accuracy. In the meantime, the generalisation capability of our framework is validated by applying the trained model on Espoo winter dataset to Espoo spring dataset and vice versa.

The remainder of this paper is organised as follows. Related work is outlined in Section II. Our novel framework is described in Section III. The experiments and the analysis are conducted in Section IV. In the end, we draw conclusions and outlook our future work.

II. Related Work

Research carried out on object recognition in point clouds can be categorised into model-driven methods and data-driven methods. Model-driven methods identify objects based on a series of manually designed rules which are constructed by empirical experience. Data-driven methods, on the contrary, recognise objects using a series of supervised statistical learning models which are trained from labelled datasets. The fundamental difference is that model-driven methods rely on handcrafted rules obtained from experiences, and data-driven methods do not require manually designed rules.

A. Model-Driven Methods

One commonly adopted model-driven method for object recognition in point clouds is the knowledge-driven method. This type of methods represents early work on object recognition in MLS data, which combine a set of manually designed rules and hand-crafted features to recognise objects in MLS data.

Techniques for recognising parameterised shapes such as planar surfaces and cylinders were reviewed in [5]. Early work to extract pole-like road furniture in MLS point clouds by introducing co-axial cylinder model fitting was presented in [1]. This method was subsequently refined in [3] by adding a scanline segmentation pre-process. A voxel-based co-axial analysis was introduced in [6] to hasten the computation compared to [1] and [3]. In [7], a percentile-based method was presented to detect pole-like road furniture in MLS data, and a set of shape templates were defined to classify traffic-functionality signs. This piece of work was optimised in [8]: rules were added with reflective feature constraints, which resulted in significant improvement in the recognition rate of traffic-functionality signs. In [9], a seed growing method was proposed to separate connected trees, and a template matching method was used to recognise pole-like furniture. A hierarchical strategy composed of rules and multi-scale supervoxels was proposed in [10] to recognise roadside objects in MLS data. In [11], a slice-cut method was proposed to identify road poles in MLS data. The method combined the co-axial cylinder fitting model and slice-wise features, which is robust to both sparse and dense point clouds.

Shape information was exploited in [12] to classify traffic signs. Compared to the work in [7], this method heavily relies

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

on the radiometric attribute of point clouds. A series of rules in combination with multi-neighbourhood features were designed in [13] to extract linear, planar and scattered elements. Two-dimensional density-based features were employed in [14] to detect poles from MLS data. These two methods are not robust to thick poles, of which eigen-based features are computed to be planar. Knowledge-driven methods are appropriate for recognition tasks with a small number of classes such as the detection of pole-like road furniture in MLS data. Numerous training samples are not required for knowledge-driven methods. This type of methods is widely applied in industrial areas because of its good interpretability.

B. Data-Driven Methods

Data-driven methods can be divided into traditional machine learning methods and deep learning methods. Traditional machine learning methods utilise handcrafted features and trained statistical learning models to identify objects. Different from traditional machine learning methods, deep learning methods automatically learn features directly from the data.

Traditional machine learning methods have been widely adopted in MLS data processing, and great progress has been achieved in recent years. In [15], Laplacian smoothing and the principal component analysis (PCA) were utilised to detect and classify pole-like objects from MLS data into three categories: utility poles, street lights and street signs. However, one piece of pole-like road furniture with multiple classes is still treated as a single one where the attached objects to poles are not separated and classified. A framework was presented in [16] to recognise objects in 3D point clouds of urban scenes based on shape and contextual features. In [17], the implicit shape models (ISM) were employed to automatically localise and recognise objects in urban scene point clouds. A template matching framework was proposed in [18] to classify urban road facilities into street signs, street lights and bus stations. The Support Vector Machine (SVM) is employed to classify pole-like road objects in MLS data [19]–[22], [36]. Random forest (RF) is adopted to semantically segment mobile laser scanning point clouds [23] and [24]. A framework was designed in [25] to classify 3D point clouds in urban road scene by employing optimal neighbourhood selection to extract features. Both [16] and [25] demonstrate that RF outperforms the other machine learning classifiers for semantic segmentation of MLS data of urban scenes.

Contextual consistency, however, is not preserved in the aforementioned frameworks which utilise simple machine learning classifiers. The work in [26] represents an early attempt on the interpretation of point clouds by utilising probabilistic graphic models. In [27], a multi-stage inference procedure was proposed to semantically segment point clouds by utilising unary classifiers, contextual features and multi-round stacking. Instead of performing inference over a graphical model, the inference procedure was taken as a composition of predictors in [28] to classify 3D point clouds. Markov Random Fields (MRFs) were adopted in [29] to recognise pole-like structures in MLS point clouds. Conditional Random Fields (CRFs) were explored in [30] to semantically segment ALS data. To address the intensive computation problem, an efficient fully connected CRF inference method was applied in [31], which was based on mean field approximation to perform fast semantic segmentation of 3D point clouds. Node features and edge features were coded in a parsimonious graphical model in [32] to semantically segment indoor scene point clouds into different categories. Multiscale features in combination with context analysis were employed in [33] to conduct semantic labelling and instance segmentation of indoor 3D point clouds. Compared to the knowledge-driven method, traditional machine learning classifiers do not require manually designed rules to perform point cloud classification.

Deep neural networks have significantly advanced point cloud interpretation in recent years [34]. However, deep learning frameworks require much more training samples than traditional machine learning classifiers. Deep learning frameworks can be categorised into three types based on the type of data fed into neural networks: voxel-based networks, multi-view projection-based networks and point-based networks.

In voxel-based networks, point clouds are structured as voxels to be fed into neural networks. An early voxel-based network was proposed in [35] to classify objects in point clouds. They voxelised point clouds into a volumetric occupancy grid, after which a supervised 3D Convolutional Neural Network (3D CNN) was introduced to classify voxels. Three-dimensional CNN was then applied in [36] to recognise objects in point clouds of urban scenes. Similarly, 3D ShapeNets was developed in [37] to recognise objects in 2.5 depth images, retrieve 3D shapes and predict the next-best-view. In [38], 3D fully convolutional neural networks (3D FCNN) and CRF were combined to semantically label point clouds. An improved VoxelNet was designed in [39] to detect objects in 3D point clouds. The OctNet was presented in [40] to perform both the semantic segmentation task and the classification task in 3D data. Voxel-based neural networks allow for hierarchical feature learning and keep multiple levels of abstraction. Therefore, they have been widely adopted in semantic segmentation, object detection and classification in point clouds.

In multi-view projection-based networks, point clouds are projected from different viewpoints to generate images. Different from voxel-based networks, multi-view projection-based networks take multi-view projected images as input for the learning process. The labelled multi-view images are then back-projected to original point clouds for prediction. In [41], a multi-view convolutional neural network was proposed for 3D shape recognition, which represents an early-stage attempt. The SnapNet was developed in [42] to semantically label 3D point clouds with 2D deep segmentation neural networks. SnapNet flexibly suits various types of point clouds such as Lidar data or photogrammetric point clouds. Similar to SnapNet, a multi-view based neural network was proposed in [43] to semantically segment 3D outdoor scenes. Because of the loss of information during the projection stage, the quality of point clouds (e.g., incomplete and noisy) heavily affects the performance of this type of methods.

Point-based neural networks directly feed networks with point clouds. Two representative pioneering point-based neural networks are PointNet [44] and PointCNN [45]. In PointNet,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

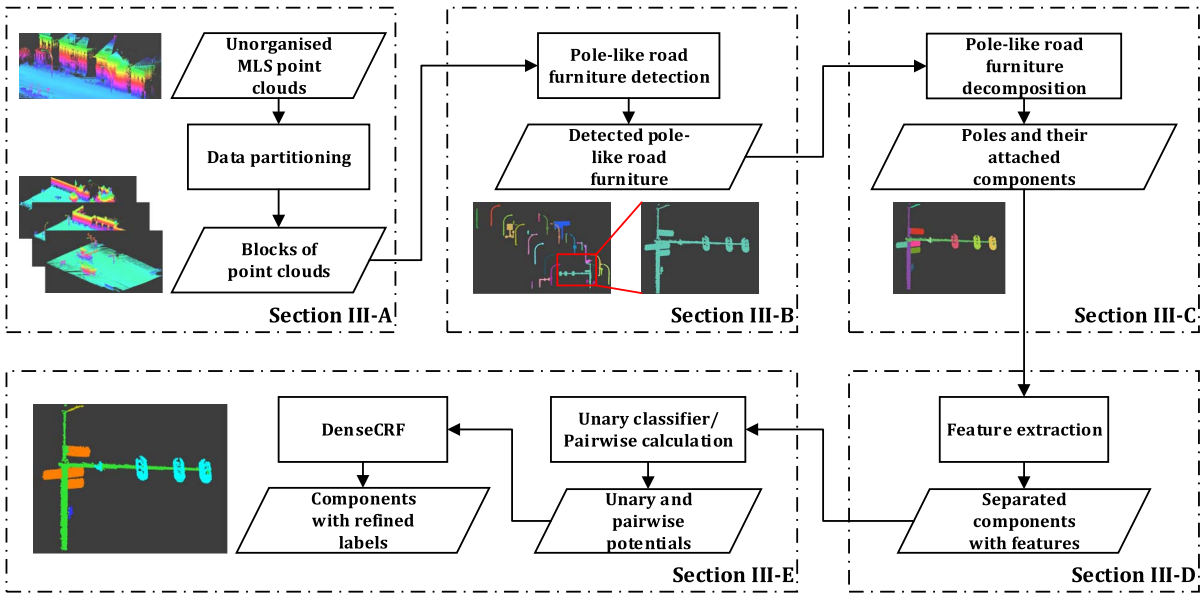


Fig. 2. The workflow of pole-like road furniture interpretation.

transformation matrices and symmetric functions were utilised to construct deep neural networks to classify objects or semantically segment scene in point clouds. PointNet, however, only preserves local features or global features and is not able to capture local structures. To tackle this problem, the PointNet++ is proposed in [46] to recognise objects in 3D point clouds by constructing a hierarchical PointNet. On the base of transformation matrices and CNN, the PointCNN was developed in [45] to semantically label point clouds. Compared to the vanilla PointNet [44], PointCNN preserves local structures in the networks. Recently, inspired by PointNet and PointCNN, many deep neural networks are architected to conduct object detection [47]–[49], semantic segmentation [50]–[54] and instance segmentation [51], [55] in 3D point clouds.

The aforementioned studies have achieved great progress for point cloud processing. However, most of them classify objects as single classes or interpret scenes at an object level. Little effort has been made on the interpretation of road scene in MLS point clouds with both semantic labels and instance information at a component level by using contextual consistency, which is also the highlight of this paper. In our research, we combine model-driven methods and data-driven methods to interpret road furniture.

III. Methodology

In this section, we introduce a framework to automatically segment point clouds of road furniture into components with both semantic and instance information, which can be divided into five stages as shown in Fig. 2: 1. unorganised point clouds are partitioned into blocks based on the recorded trajectory information, 2. pole-like road furniture is extracted from each block of point clouds, 3. the pieces of furniture are segmented into poles and their attached components, 4. features are extracted from these separated components, 5. components are classified based on the extracted features using a combination of a simple classifier and DenseCRF.

In the fifth stage, unary potentials are obtained from the trained classifier to depict probabilities of classes for the components. Pairwise potentials are calculated encoding contextual consistency and fed into the constructed DenseCRF which refine the intermediate classification from the simple unary classifiers.

A. Data Partitioning

To prevent memory overflow caused by the large size of MLS data and reduce computation time, we first divide the data into blocks along the trajectory of the mobile vehicle both on straight roads (Fig. 3a) and curved roads (Fig. 3b). The trajectory is split into line segments with a manually specified length, and then the whole point cloud is sliced into blocks perpendicular to the trajectory at the segment edges.

B. Pole-Like Road Furniture Detection

In this stage, ground points are first removed from each block. A point is regarded as a ground point if it has a small standard deviation of height variance (SDHV) of its neighbouring points and below a relative elevation to its corresponding trajectory point. The threshold of SDHV is set to be 0.15m, and the threshold of the relative elevation is decided by the mounting height of laser scanners.

Then pole-like road furniture is detected from the remaining above-ground points. A connected component analysis is performed to segment above-ground points into separate objects. A rough classification is then conducted to identify the objects as buildings, trees or pole-like road furniture. An object is recognised as a building according to the height and 2D length of the bounding box and the area of the extracted surface, since both the bounding box of a building and the area of a

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

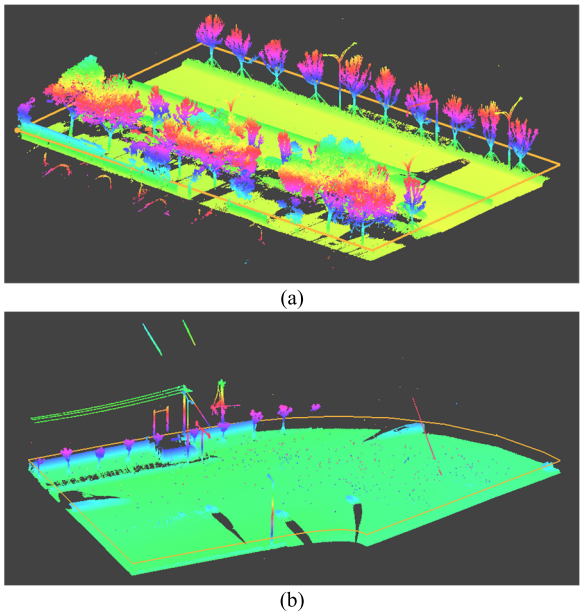


Fig. 3. Two partitioned blocks of point clouds coloured by their elevation (Yellow lines represent block lines). (a) One straight road segment. (b) One curved road segment.

building façade are supposed to be large. Trees are recognized and removed based on the rule that trees are much lower than other objects in terms of the ratio of points with the first return. In the meantime, pole-like road furniture is detected using a slice-cutting algorithm combined with a co-axial cylinder fitting analysis, and an occlusion analysis is implemented to exclude incorrectly detected pole-like objects behind building façades [4].

C. Pole-Like Road Furniture Decomposition

A decomposition algorithm is adopted to segment the detected pole-like road furniture into poles and their attached components (e.g., street light heads, traffic signs, etc.) [56].

Three methods are implemented to extract poles from pole-like road furniture based on their types. The 2D density-based method is designed for the furniture with many components. The random sample consensus (RANSAC) line fitting method is proposed to cope with the furniture consisting of horizontal poles. The slice-cutting based method is used for the remainder.

Components attached to poles are separated into individual components based on their spatial relations after removing the extracted poles. A connected component analysis is first used, and a set of splitting and merging rules are then utilised to refine the over-segmentation and under-segmentation of components [4].

D. Feature Extraction

As each segmented component represents a meaningful part of road furniture, we treat each component as a unit to extract its features. Based on the composition of features, we divide features into two groups: unary features and contextual features as shown in Fig. 4. Compared to other methods

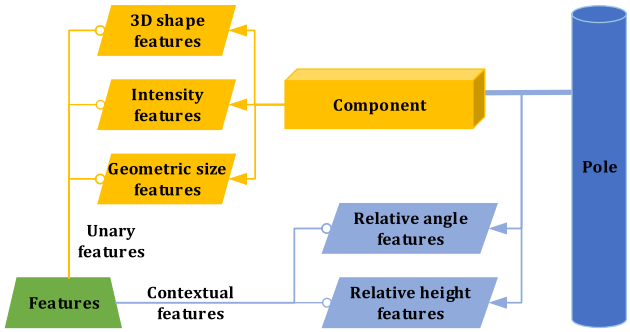


Fig. 4. Component-wise features.

using traditional pointwise features or voxel-wise features, our method extracts component-wise features and thus preserves more important clues such as the relation between different components.

1) Unary Features: Three sets of unary features are extracted directly from the components: 3D shape features for describing shapes, reflectance features for material properties, and geometric features for geometric attributes such as size.

3D shape features are obtained from the eigenvalues λ_i of the covariance matrix of the point cloud of a component, including the normalised eigenvalues e_i (Eq. 1), linearity E_L, planarity E_P, scattering E_S, linear-planarity E_{LP}, omnivariance E_O, anisotropy E_A and entropy E_E (Equations 2 to 8).

$$e_i = \lambda_i \Big/ \sum\nolimits_i \lambda_i, \quad i \in \{1,2,3\}, \ \lambda_1 > \lambda_2 > \lambda_3 \geq 0 \tag{1}$$

$$E_L = (e_1 - e_2) \big/ e_1 \tag{2}$$

$$E_P = (e_2 - e_3) \big/ e_1 \tag{3}$$

$$E_S = e_3 \big/ e_1 \tag{4}$$

$$E_{LP} = (e_2 - e_3) \big/ e_2 \tag{5}$$

$$E_O = \sqrt[3]{e_1 e_2 e_3} \tag{6}$$

$$E_A = (e_1 - e_3) \big/ e_1 \tag{7}$$

$$E_E = \sum\nolimits_{i=1}^{3} e_i \ln(e_i) \tag{8}$$

The two-dimensional reflectance features are designed to indicate the intensity attribute of components which are used for the recognition of traffic-functionality signs. They are composed of the median value I_M and the average value I_A of the recorded intensity of points for each component.

Geometric features are extracted by constructing a bounding box for each component. The features include the length G_L, the width G_W, the height G_H, the 2D maximum range G_R which is the 2D largest distance on the XY plane between two points in a component, the volume G_V and the ratio feature G_{RH}. The volume feature and the ratio feature are calculated as follows:

$$G_V = G_L \cdot G_W \cdot G_H \tag{9}$$

$$G_{RH} = G_R \big/ G_H \tag{10}$$

2) Contextual Feature: Different from unary features which depict attributes of components themselves, contextual features describe the relative position between poles and their attached

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

components. Contextual features are composed of relative height features and relative angle features.

Relative height features are the height differences between the key positions of components and poles, where the key positions include the lowest (sub-indexed by L), the centre (sub-indexed by C) and the highest (sub-indexed by H). There are five relative height features [2] including RH_{CL}, RH_{CH}, RH_{LL}, RH_{HL}, RH_{HH}, where the first sub-index of each notation denotes the key position of a component, and the second denotes the key position of a pole.

Relative angle features are angles between poles and their attached components. Besides the two features RA_{NP}, RA_{NV} proposed in our previous work [2], we add two more features RA_{PV} and RA_{PP} in this paper. The four features are described in the following equations:

$$RA_{NP} = \left|\overrightarrow{C_N} \cdot \overrightarrow{P_P}\right| = \left|\cos \alpha_{NP}\right| \tag{11}$$

$$RA_{NV} = \left|\overrightarrow{C_N} \cdot \overrightarrow{V_V}\right| = \left|\cos \alpha_{NV}\right| \tag{12}$$

$$RA_{PV} = \left|\overrightarrow{C_P} \cdot \overrightarrow{V_V}\right| = \left|\cos \alpha_{PV}\right| \tag{13}$$

$$RA_{PP} = \left|\overrightarrow{C_P} \cdot \overrightarrow{P_P}\right| = \left|\cos \alpha_{PP}\right| \tag{14}$$

where $\overrightarrow{C_N}$ and $\overrightarrow{C_P}$ are the unit vectors of the normal direction and the principal direction of a component respectively, and $\overrightarrow{P_P}$ and $\overrightarrow{V_V}$ are the unit vectors of the principal direction and vertical direction of the connected pole of the component. α_{NP} is the relative angle between $\overrightarrow{C_N}$ and $\overrightarrow{P_P}$, α_{NV} is the angle between $\overrightarrow{C_N}$ and $\overrightarrow{V_V}$, α_{PV} is the angle between $\overrightarrow{C_P}$ and $\overrightarrow{V_V}$, and α_{PP} is the angle between $\overrightarrow{C_P}$ and $\overrightarrow{P_P}$.

E. Probabilistic Component Classification

Features extracted in Section III-D are utilised as the input for the classification of the components by using DenseCRF combined with a classifier. DenseCRF [57] is employed to preserve contextual consistency, i.e., to model the relations between different components. In contrast to traditional CRF, DenseCRF is capable of encoding potentials for long-range relations (i.e., relations between not only adjacent nodes but also non-adjacent nodes). Thus, DenseCRF is well suited in our case to learn the mounting rules of components in urban scenes, since in our research, plenty of components in the same category share long-range relations. For instance, the distance between neighbouring street lights is often fixed. By using DenseCRF, more relations between components can be encoded and more contextual consistency patterns could be preserved.

1) Definition of DenseCRF: Our DenseCRF model is essentially an undirected graph $g = \{\nu(S), \varepsilon(S)\}$ (Fig. 5). Here $S = \{S_1, \ldots, S_N\}$ denotes the N components extracted in the decomposition stage. $\nu(S)$ is the set of nodes in the graph which are fully connected, and each node represents an extracted component S_i. $\varepsilon(S)$ is the set of undirected edges between different nodes in the graph.

We further define the class labels of the components as $I = \{I_1, \ldots, I_N\}$, where the domain of each I_i is a set of class labels $L = \{l_1, \ldots, l_k\}$. A random field $(I|S, \omega)$ conditioned on S and model parameters $\omega = \{\omega_\mu, \omega_p\}$ is characterized by a Gibbs distribution as follows [57]:

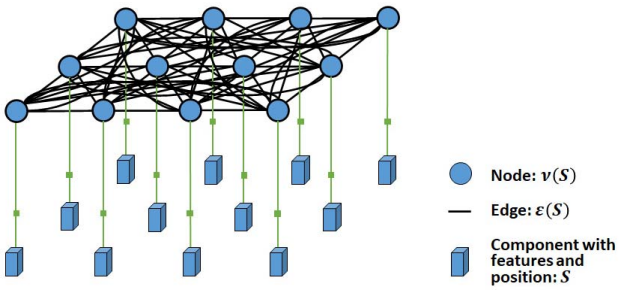


Fig. 5. The constructed DenseCRF.

$$P(I|S, \omega) = \exp\left(-E(I|S, \omega)\right) / Z(S, \omega) \tag{15}$$

where the marginal partition function $Z(S, \omega)$ transforms the potentials into probabilities. The Gibbs energy function $E(I|S, \omega)$ of a label assignment is defined as follows:

$$E(I|S, \omega) = \sum_i \varphi_\mu\left(I_i|S_i; \omega_\mu\right) + \sum_{i=1}^{N} \sum_{j=1}^{i} \varphi_p\left(\left(I_i, I_j\right)|\left(S_i, S_j\right); \omega_p\right) \tag{16}$$

where φ_μ is the unary potential parameterised by ω_μ, and φ_p is the pairwise potential parameterised by ω_p.

2) Unary Potential: The unary potential φ_μ for a component is the probability of the component being assigned to a specific class label l_m which is described as follows

$$\varphi_\mu\left(I_i = l_m|S_i; \omega_\mu\right) = P\left(I_i = l_m|S_i; \omega_\mu\right) \tag{17}$$

A classifier, which takes the unary features and contextual features of the components as the input, is adopted to compute φ_μ for each class of every component. In this paper, six different classifiers are tested to compare their performances including a knowledge-driven method, three discriminative learning models, random forest (RF), support vector machine (SVM) and multinomial logistic regression (MLR), and two generative learning models, Gaussian mixture model (GMM) and naïve Bayes (NB).

For the knowledge-driven method, we define a set of rules to classify these components empirically learned from the mounting regulations of road furniture. These rules are encoded as a set of feature constraints formulated for the corresponding template of each class. Take traffic signs as an instance, they are constrained by the following features: 1. Traffic signs are normally connected to vertical poles, and the connected position is not assumed to be located in the lowest part of their attached poles. 2. Traffic signs contain parts that are close to planar and have high intensity in the point cloud. 3. The relative angle between the normal direction of a traffic sign and the principal direction of its connected pole is perpendicular. Detailed description can be found from our previous work [58].

As one of the most superior traditional machine learning classifiers, RF has been widely applied in the classification of point clouds [24]. In this paper, the RF classifier ensembles the mounting rules of components in urban road scenes with

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

numerous weak learners [59] to predict the probability of each class for every component. We feed RF manual labels of components and their extracted features υ for training. Trees are constructed for the data augmented by the bagging strategy. To explore the effectiveness of our designed features, the Gini index is introduced to compute the feature importance as in [60]. Two parameters, the number of trees N_T and the depth of trees D_T, are utilised and tuned in the RF classifier, and a grid search is used to obtain the optimal combination of these two parameters.

SVM is a binary classifier that maximizes the margin, that is, the distance of the closest vectors to the decision surface in both classes [61]. We use the one-vs-all strategy to solve the multiclass problem. Radial basis function (RBF) kernel is used with the C-SVM to enable nonlinear decision surfaces. The model contains hyperparameters γ and C for the training and prediction. Similar to RF, we use a grid search for hyperparameter tuning. To mitigate effects from the unimportant features, the backward elimination [62] is utilised for feature selection.

MLR is to build a linear predictor function with a set of linearly combined weights to predict probabilities of observations.

GMM in this paper assumes that all the components with features are generated from the mixture of a certain number of Gaussian distributions with unknown parameters. In other words, GMM is a sum of multivariate Gaussian distributions, and each distribution represents one class of components. In our implementation, components with features and manual labels are fed into GMM to estimate the parameters using expectation maximization (EM).

NB assumes that given the class label, the value of a particular feature is independent of the values of any other features, and each value of the features follows a Gaussian distribution. The maximum a posterior (MAP) is adopted to estimate the parameters.

3) Pairwise Potential: Pairwise potential is designed to explore the contextual relations between different components during the classification. The pairwise potential φ_p is encoded by two kernels, the appearance kernel k_{app} and the relative location kernel k_{relloc} expressed in Equations 18-20. The appearance kernel is to emphasize that nearby components sharing similar features are likely to be in the same class based on the observation, while the relative location kernel is to encourage components with a similar height to be in the same class.

$$\begin{aligned}\varphi_p\left(\left(I_i, I_j\right)\middle|\left(S_i, S_j\right); \omega_p\right) &= \mu\left(I_i, I_j\right) k\left(\left(S_i, S_j\right); \omega_p\right) \\ &= \mu\left(I_i, I_j\right)\left\{\omega_p^{(1)} k_{\text{app}}\left(S_i, S_j\right)\right. \\ &\quad \left. + \omega_p^{(2)} k_{\text{relloc}}\left(S_i, S_j\right)\right\} \qquad (18)\end{aligned}$$

$$\begin{aligned}k_{\text{app}}\left(S_i, S_j\right) &= \exp\left(-\left(\left\|p_i - p_j\right\| - d\right)^2 \middle/ 2\theta_\alpha^2\right. \\ &\quad \left. - \left|f_i - f_j\right|^2 \middle/ 2\theta_\beta^2\right) \qquad (19)\end{aligned}$$

$$k_{\text{relloc}}\left(S_i, S_j\right) = \exp\left(-\left(\left\|p_i - p_j\right\| - d\right)^2 \middle/ 2\theta_\gamma^2\right) \qquad (20)$$

where p_i and p_j are the positions of components S_i and S_j, and f_i and f_j are the feature vectors of S_i and S_j. $\mu\left(I_i, I_j\right)$ is a simple label compatibility function given by the Potts model which penalises class changes with similar components. The kernel parameters $\theta = \left\{\theta_\alpha, \theta_\beta, \theta_\gamma\right\}$ are utilised to control the degree of proximity and similarity. $\omega_p^{(1)}$ and $\omega_p^{(2)}$ are the pairwise parameters of the potential.

4) Inference and Training: For the inference, the KL-divergence is adopted in our model in combination with mean fields to approximate the probability distribution function [57] instead of computing the CRF distribution directly using Markov chain Monte Carlo (MCMC). It is because that it is difficult to make the model converge efficiently since the computational complexity of message passing between nodes is extremely high in the graph due to the large number of edges in DenseCRF. In the message passing process, we use the high-dimensional filtering convolution with a Gaussian kernel. This reduces the complexity of the intractable message passing process to linear, which highly facilitates the inference process.

For the training, the intersection over union (IOU) is used as the loss function which mitigates imbalances in the training data, and gradient-based optimisation is used to minimise the loss function. The gradient-based optimisation includes the mean field gradient, unary and pairwise gradients. The label compatibility and kernel parameters are learned with the tuning of unary and pairwise gradients. The parameters of unary potentials are learned during the training of the unary classifiers. With the trained model, the label of a component is predicted as the class with the highest probability:

$$\tilde{I} = \arg\max_{I \in \mathcal{L}} P\left(I | S; \omega\right) \qquad (21)$$

IV. Experimental Results

To evaluate the performance of our framework, we conduct experiments on three datasets, one collected in Enschede, a medium-sized city in the east of the Netherland, and the other two in Espoo, the second-largest city of Finland, in different seasons. Detailed description of our datasets can be found in our previous work [2]. Our experiments are only conducted on the three datasets since there are a considerable number and types of pole-like road furniture in these three datasets, whereas the number of pole-like road furniture samples in other benchmark datasets is limited. Besides, the appearance of pole-like road furniture is noticeably different between Enschede and Espoo.

A. Results

To quantitively evaluate the reliability of our framework, we compare it to our previous work [2] in which merely unary classifiers were employed. The reliability on different datasets has been thoroughly validated by the remarkable performance of our framework as shown in Table I. The highest overall accuracy (OA) of classification of components is above 83.5% in the Enschede dataset, and above 95.0% in both Espoo winter and spring dataset.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
THE OVERALL ACCURACY OF ROAD FURNITURE INTERPRETATION OF FIVE DIFFERENT METHODS

Method \ Dataset	Enschede	Espoo winter	Espoo spring
KD	64.4%	51.5%	59.5%
KD-DenseCRF	72.7%	73.8%	86.5%
SVM	72.2%	79.1%	78.2%
SVM-DenseCRF	76.7%	85.1%	89.6%
RF	81.0%	92.3%	94.1%
RF-DenseCRF	83.7%	96.4%	95.3%
MLR	21.4%	19.3%	14.1%
MLR-DenseCRF	45.1%	57.4%	51.4%
GMM	29.0%	35.4%	40.7%
GMM-DenseCRF	59.6%	61.9%	85.7%
NB	59.1%	86.8%	86.8%
NB-DenseCRF	65.5%	90.9%	90.0%

Then the combinations of every single unary classifier and DenseCRF are fully explored as shown in Table I. Amongst all these methods, the combination of RF and DenseCRF (RF-DenseCRF) performs the best, and DenseCRF improves the accuracy of RF, the state-of-the-art method in the interpretation of road furniture [2], by 2.7%, 4.1% and 1.2% in Enschede, Espoo winter and Espoo spring datasets respectively. This may be due to that RF takes the advantage of the combination of numerous weak learners, and DenseCRF works as a smoother to reallocate the predictions. The MLR classifier achieves the lowest accuracy which implies that the classification problem is nonlinear as MLR utilises a simple linear function for prediction. Besides, the significant difference between the performance of our framework on the Enschede dataset and the two Espoo datasets is because that more classes and more complex road furniture are included in the Enschede dataset which makes the classification more challenging.

We further generate confusion matrices of the results on Enschede, Espoo winter and Espoo spring datasets as illustrated in Tables II, III, and IV respectively to evaluate the combination of RF and DenseCRF. The F-scores of the recognition of street light in all three datasets is extremely high ($>$ 96.0%). In Espoo winter and spring datasets, our framework achieves a high F-score of above 95.0% for the recognition of traffic signs and other attachments and a much lower F-score for street signs. The low score might be due to that it is difficult for the classifier to retrieve enough representative information since there are few training and testing street sign samples. The same cases also arise in the recognition of traffic lights and other signs.

There is a remarkable improvement for the recognition F-score of traffic signs in all three datasets. The reason is that DenseCRF is able to capture mounting patterns of traffic signs when there are sufficient training samples. By contrast, the recognition F-score of traffic lights in Enschede dataset decreases significantly due to the small number of traffic lights in the Enschede training and testing datasets. This implies that the mounting patterns can hardly be learned. Similar cases occur in the recognition of street signs on the Espoo spring dataset.

The visualisation of our results on the three datasets is shown in Fig. 6a-c. In the Enschede dataset, most street lights (colourised to be yellow) are well recognised. Traffic signs and street signs (colourised to be magenta and orange) are also decently identified (Fig. 6a). A few components are incorrectly recognised as other objects. For instance, street lights and traffic lights are misclassified as other objects (the bottom middle and the bottom right subfigures in Fig. 6a). This case of misclassification is aroused by the incorrect decomposition. In Espoo winter (Fig. 6b) and spring datasets (Fig. 6c), our framework achieves great performance. Components of classes amongst street light, traffic signs and other attachments in these two datasets are mostly predicted with the correct labels. Because of the lack of training and testing samples with street signs in these two datasets, three other objects are misclassified as traffic signs (the bottom left and bottom right subfigures in Fig 6c). Besides these negative results, the high similarity between powerline fragments and street light heads lead to powerline segments misrecognised as street lights (the bottom middle subfigure in Fig. 6b). These two classes are elongated, and their principal directions are perpendicular to the principal direction of their attached poles.

To test the reliability of our framework, we compare the performance of our framework on Espoo winter and spring datasets collected in two different epochs in nearly the same area. Table II and Table III suggest that the F-scores of street light, traffic sign and other attachments are all above 95%, and the F-score of traffic sign is between 70% and 75%. The consistency of the performances between these two datasets demonstrates that our framework is capable of dealing with different datasets composed of the same type of road furniture with consistently good performance.

Finally, the generalisation capability of our framework is assessed. To validate the generalisation capability of our trained model, the model trained on the Espoo spring training data is tested on the Espoo winter testing data with an OA of 93.1% (Table V), and the model trained on the Espoo winter training data is tested on the Espoo spring testing data with an OA of 95.1% (Table VI). The high OA ($>$ 93%) of both tests indicate a high generalisation capability of our framework. One exception is that the F-score of traffic sign in the Espoo winter dataset and spring dataset is lower compared to the result predicted by their original trained model. This is because the point clouds of traffic signs in the Espoo spring dataset are surrounded with stray points in the air more than in the Espoo winter dataset, which elevate the importance of intensity features. The much higher intensity in the Espoo spring dataset gives rise to a much lower F-score of traffic signs in Espoo winter testing data predicted by the Espoo spring trained model.

B. Discussion

We further investigate the advantages of RF-DenseCRF. As illustrated in Section IV-B, the RF-DenseCRF outperforms

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE II
THE CONFUSION MATRIX OF RESULTS IN THE ENSCHEDE DATASET

Class	Street light	Traffic sign	Street sign	Traffic light	Other signs	Other attachments	Correctness
Street light	148	0	0	2	0	0	98.7%
Traffic sign	0	73	13	1	11	1	73.7%
Street sign	0	5	42	0	1	0	87.5%
Traffic light	0	1	0	14	6	3	58.3%
Other signs	0	4	0	1	13	5	56.5%
Other attachments	1	3	1	5	8	81	81.8%
Completeness	99.3%	84.9%	75.0%	60.9%	33.3%	90.0%	
F-score	99.0%	78.9%	80.8%	59.6%	41.9%	85.7%	
F-score of previous work	99.0%	64.7%	76.6%	70.8%	40.0%	86.1%	

TABLE III
THE CONFUSION MATRIX OF RESULTS IN THE ESPOO WINTER DATASET

Class	Street light	Traffic sign	Street sign	Other attachments	Correctness
Street light	91	0	0	4	95.8%
Traffic sign	0	108	1	2	97.3%
Street sign	0	1	4	1	66.7%
Other attachments	1	3	0	146	97.3%
Completeness	98.9%	96.4%	80.0%	95.4%	
F-score	97.3%	96.8%	72.7%	96.3%	
F-score of previous work	96.2%	90.7%	61.5%	94.8%	

TABLE IV
THE CONFUSION MATRIX OF RESULTS IN THE ESPOO SPRING DATASET

Class	Street light	Traffic sign	Street sign	Other attachments	Correctness
Street light	114	0	0	2	98.3%
Traffic sign	0	129	4	0	97.0%
Street sign	0	0	12	3	80.0%
Other attachments	7	6	1	212	93.8%
Completeness	94.2%	95.6%	70.6%	97.7%	
F-score	96.2%	96.3%	75.0%	95.7%	
F-score of previous work	95.6%	93.4%	81.3%	96.0%	

TABLE V
THE ESPOO SPRING TRAINED MODEL (RF-DENSECRF) ON THE ESPOO WINTER TESTING DATA

Class	Street light	Traffic sign	Street sign	Other attachments	Correctness
Street light	92	0	0	4	95.8%
Traffic sign	0	93	0	2	97.9%
Street sign	0	4	5	0	55.6%
Other attachments	0	15	0	147	90.7%
Completeness	100%	83.0%	100%	96.1%	93.1%(OA)
F-score	97.9%	89.9%	71.4%	93.3%	

TABLE VI
THE ESPOO WINTER TRAINED MODEL (RF-DENSECRF) ON THE ESPOO SPRING TESTING DATA

Class	Street light	Traffic sign	Street sign	Other attachments	Correctness
Street light	112	1	0	2	97.4%
Traffic sign	1	130	5	3	93.5%
Street sign	0	0	12	0	100.0%
Other attachments	8	4	0	212	94.6%
Completeness	92.6%	96.3%	70.6%	97.7%	95.1% (OA)
F-score	94.9%	94.9%	82.8%	96.1%	

the other methods. Random forest is an ensemble learning classifier consisting of numerous weak learners, where each learner behaves as an empirically defined rule latently contained in our hand-crafted features. For instance, to enable drivers to easily catch sight of traffic information, street signs should not be mounted too high or too low. It is easy for a weak learner in RF to capture this type of mounting pattern. In contrast, the other classifiers heavily rely on hyperplane construction or distributions (e.g., Gaussian), which are not prominent amongst our designed features. Even the infinite-dimensional feature space by the kernel trick in SVM is not sufficient to describe mounting patterns of road furniture. Therefore, RF outperforms the other methods by making better use of our designed features. In the meantime, DenseCRF can help to find patterns between different components, and improve the performance by learning such patterns with the pairwise potentials. For example, neighbouring street lights

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

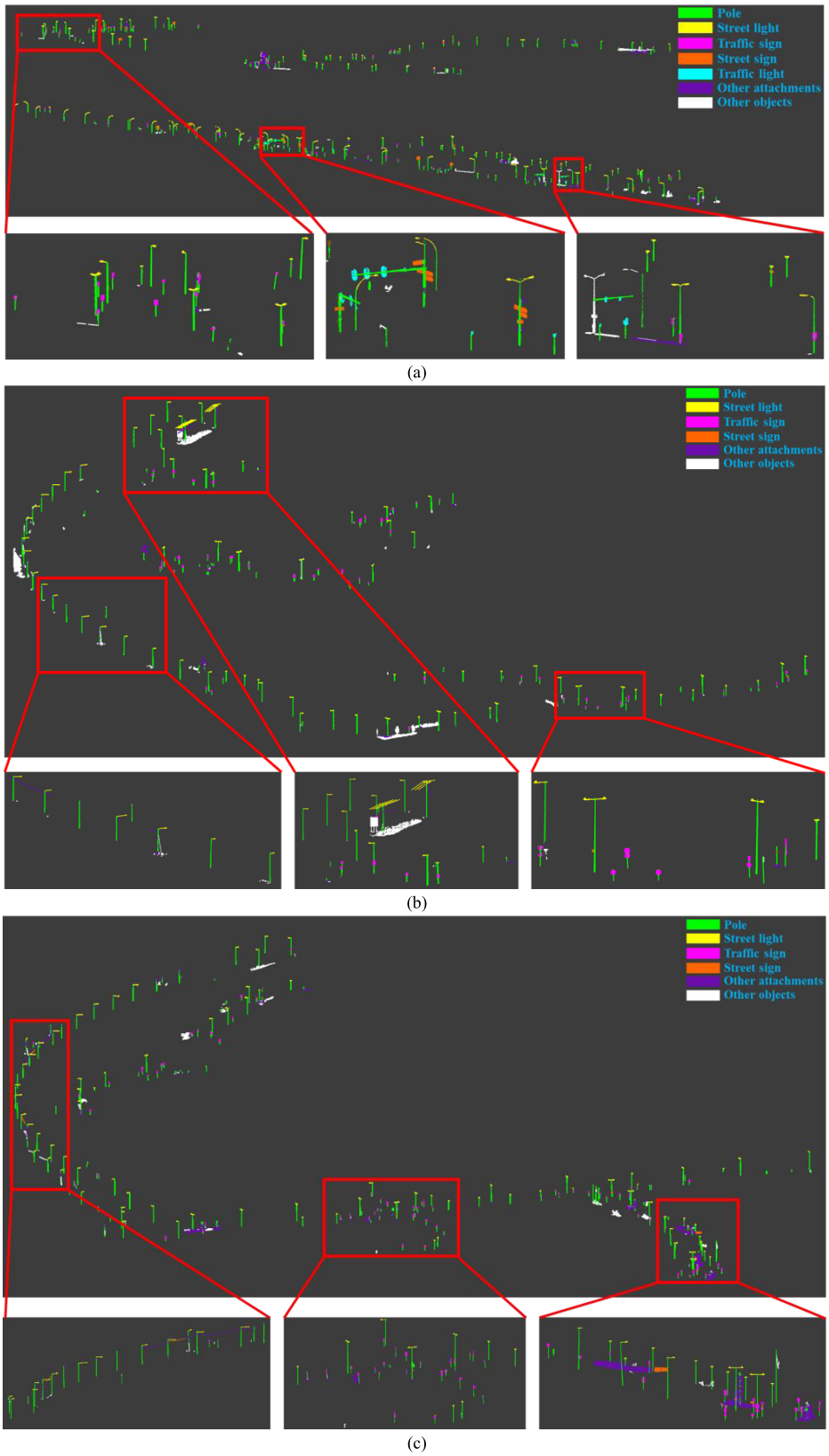


Fig. 6. The visualisation of instance-aware semantic segmentation of road furniture in (a) Enschede dataset, (b) Espoo winter dataset and (c) Espoo spring dataset.

are usually mounted with a fixed distance. And DenseCRF is capable of capturing this contextual consistency information. However, when there is only a small number of training or testing samples, it is rather challenging for DenseCRF to improve the classification of components, or the performance may even go down.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE VII
THE 12 MOST IMPORTANT FEATURES, RANKED BY FEATURE IMPORTANCE

Rank	Feature		
	Enschede	Espoo winter	Espoo spring
1	RH_{HH}	RH_{CL}	I_M
2	RH_{LL}	G_V	RA_{PP}
3	RA_{PP}	e_2	I_A
4	G_R	e_1	G_W
5	RA_{PV}	RA_{PV}	RH_{CH}
6	E_L	G_R	RA_{PV}
7	E_{LP}	RA_{NP}	E_{LP}
8	G_W	RH_{CH}	RH_{CH}
9	I_M	I_M	G_V
10	E_O	E_{LP}	G_R
11	e_1	RA_{NV}	e_2
12	G_L	G_W	E_L

Then we explore the importance of our designed features. The importance of the unary and contextual features empirically designed based on the mounting rules of road furniture is calculated by the Gini index following [60] as shown in Table VII. Contextual features $\{RH, RA\}$ are the three most important features during the training of the Enschede dataset. This is because contextual features are prominent for differentiating components. For instance, it is difficult to distinguish street light and traffic light only by using a unary feature, but much easier to distinguish them by using the contextual feature RA_{PP}. The median intensity I_M in the training of the Enschede dataset and the Espoo winter dataset is not as important as in the training of Espoo spring dataset. I_M is designed for recognising traffic-functionality signs and other attachments. In the training of the Enschede dataset, eigenvalue features (E_L and e_2) are already enough to differentiate most of them. Compared to the Enschede dataset and Espoo winter dataset collected with good quality, there are stray points in the Espoo spring dataset, which makes the eigenvalue features unreliable. Therefore, I_M takes over the role of eigenvalue features for differentiating traffic signs and other components. This is also the reason why eigenvalue features are more important in the training of the Espoo winter dataset than in the training of the Espoo spring dataset.

C. Ablation Study

To validate the effectiveness of the DenseCRF, we carry out an ablation study and compare our research to the other work [10] as illustrated in Fig. 7. Our framework outperforms [10], which significantly advances the instance-aware segmentation of road furniture in MLS data. The ablation study is performed between RF and RF-DenseCRF. As shown in Fig. 7, the contextual consistency in DenseCRF improves the performance of the RF method.

In addition, the computation cost is recorded with the training and testing period (Table VIII). All the training and inference time is less than 7s with our framework, which indicates the high efficiency of our framework. The configuration of our computation platform is Intel i7-8700k (6-core), 64G RAM and Nvidia RTX 2070 video card. The pairwise and kernel parameters of our DenseCRF model are automatically learned in the training process.

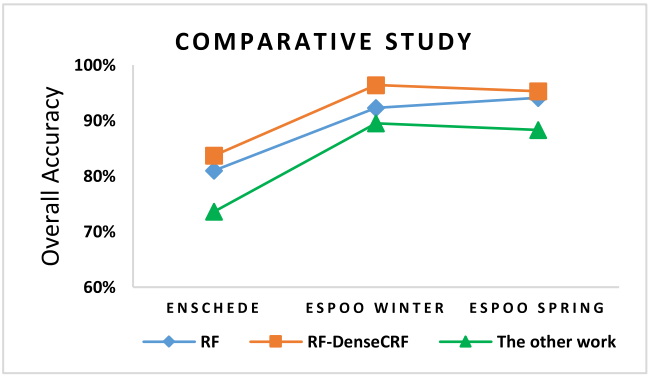


Fig. 7. The performance comparison amongst RF, RF-DenseCRF and [10].

TABLE VIII
THE COMPUTATION TIME OF OUR FRAMEWORK

Computation time (ms)	RF training	RF testing	DenseCRF training	DenseCRF testing
Enschede	81.6	19.1	5972.4	12.0
Espoo winter	66.4	14.8	2211.8	9.2
Espoo spring	54.4	19.5	2416.8	11.5

V. CONCLUSION

A novel framework for automatic recognition of road furniture is proposed in this paper which constructs a classification module combining a simple classifier and DenseCRF, and the state-of-the-art performance is thus achieved for the instance-aware semantic segmentation of road furniture in mobile scanning data. New contextual features are designed and incorporated into the previous features for the calculation of unary potentials used in DenseCRF to learn the mounting patterns of pieces of road furniture. Meanwhile, pairwise potentials are calculated and used in DenseCRF to preserve contextual consistency between components of road furniture.

Our framework performs well on three datasets in Enschede and Espoo test sites which are in two different countries. By using DenseCRF, we elevate the accuracy of the Enschede dataset from 81.0% to 83.7% and the accuracy of Espoo winter and spring datasets from >92% to >95% compared to our previous work [2]. Besides, combinations of six different classifiers with and without DenseCRF are carefully evaluated, and RF-DenseCRF performs the best. By our designed framework, we can robustly interpret road furniture with both semantic label and instance information based on their functionalities.

Furthermore, only small differences exist in the performance of our framework between the two epochs (winter and spring) in Espoo, which proves that our designed framework is generic, reliable, and robust. Our framework thus shows great potential for creating 3D high-definition (HD) maps, which are crucial for autonomous driving and urban road inventory.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

In the future work, we will combine point clouds and images to deal with the segmentation of urban roadside objects. Currently, mobile laser scanning point clouds can only provide the information of structural features and single-channel intensity. Our algorithm is therefore restrained by the limited information which is not sufficient for a more detailed road furniture interpretation. For instance, it is difficult to recognise the speed limit in a sign by merely using point clouds. Instead, combining with the information provided by images, we believe that a better interpretation of road furniture can be achieved.

REFERENCES

[1] C. Brenner, "Extraction of features from mobile laser scanning data for future driver assistance systems," in *Proc. 12th Agile Conf.*, Hannover, Germany, 2009, pp. 25–42.

[2] F. Li *et al.*, "Semantic segmentation of road furniture in mobile laser scanning data," *ISPRS J. Photogramm. Remote Sens.*, vol. 154, pp. 98–113, Aug. 2019.

[3] M. Lehtomäki, A. Jaakkola, J. Hyyppä, A. Kukko, and H. Kaartinen, "Detection of vertical pole-like objects in a road environment using vehicle-based laser scanning data," *Remote Sens.*, vol. 2, no. 3, pp. 641–664, 2010.

[4] F. Li, S. O. Elberink, and G. Vosselman, "Pole-like road furniture detection and decomposition in mobile laser scanning data based on spatial relations," *Remote Sens.*, vol. 10, no. 4, p. 531, 2018.

[5] G. Vosselman, B. G. H. Gorte, G. Sithole, and T. Rabbani, "Recognising structure in laser scanner point clouds," *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 46, no. 8, pp. 33–38, 2004.

[6] C. Cabo, C. Ordoñez, S. García-Cortés, and J. Martínez, "An algorithm for automatic detection of pole-like street furniture objects from mobile laser scanner point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 47–56, Jan. 2014.

[7] S. Pu, M. Rutzinger, G. Vosselman, and S. O. Elberink, "Recognizing basic structures from mobile laser scanning data for road inventory studies," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 6, pp. S28–S39, Dec. 2011.

[8] D. Li and S. O. Elberink, "Optimizing detection of road furniture (pole-like object) in mobile laser scanner data," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 2, no. 5, pp. 163–168, 2013.

[9] S. O. Elberink and B. Kemboi, "User-assisted object detection by segment based similarity measures in mobile laser scanner data," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XL-3, pp. 239–246, Aug. 2014.

[10] B. Yang, Z. Dong, G. Zhao, and W. Dai, "Hierarchical extraction of urban objects from mobile laser scanning data," *ISPRS J. Photogram. Remote Sens.*, vol. 99, pp. 45–57, Jan. 2015.

[11] F. Li *et al.*, "Pole-like road furniture detection in sparse and unevenly distributed mobile laser scanning data," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 4, no. 2, pp. 1–8, 2018.

[12] B. Riveiro, L. Díaz-Vilariño, B. Conde-Carnero, M. Soilán, and P. Arias, "Automatic segmentation and shape-based classification of retro-reflective traffic signs from mobile LiDAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 1, pp. 295–303, Jan. 2016.

[13] M. Bremer, V. Wichmann, and M. Rutzinger, "Eigenvalue and graph-based object extraction from mobile laser scanning point clouds," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. II-5/W2, pp. 55–60, Oct. 2013.

[14] S. I. El-Halawany and D. D. Lichti, "Detecting road poles from mobile terrestrial laser scanning data," *Sci. Remote Sens.*, vol. 50, no. 6, pp. 704–722, Dec. 2013.

[15] H. Yokoyama, H. Date, S. Kanai, and H. Takeda, "Detection and classification of pole-like objects from mobile laser scanning data of urban environments," *Int. J. CAD/CAM*, vol. 13, no. 2, pp. 1–10, 2013.

[16] A. Golovinskiy, V. G. Kim, and T. Funkhouser, "Shape-based recognition of 3D point clouds in urban environments," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep. 2009, pp. 2154–2161.

[17] A. Velizhev, R. Shapovalov, and K. Schindler, "Implicit shape models for object detection in 3D point clouds," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. I-3, pp. 179–184, Jul. 2012.

[18] Y. Yu, J. Li, H. Guan, and C. Wang, "Automated extraction of urban road facilities using mobile laser scanning data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2167–2181, Aug. 2015.

[19] B. Yang and Z. Dong, "A shape-based segmentation method for mobile laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 81, pp. 19–30, Jul. 2013.

[20] M. Lehtomäki *et. al*, "Object classification and recognition from mobile laser scanning point clouds in a road environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 1226–1239, Feb. 2016.

[21] M. Soilán, B. Riveiro, J. Martínez-Sánchez, and P. Arias, "Traffic sign detection in MLS acquired point clouds for geometric and image-based semantic inventory," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 92–101, Apr. 2016.

[22] J. Wang, R. Lindenbergh, and M. Menenti, "SigVox—A 3D feature matching algorithm for automatic street object recognition in mobile laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 128, pp. 111–129, Jun. 2017.

[23] K. Fukano and H. Masuda, "Detection and classification of pole-like objects from mobile mapping data," *ISPRS Ann. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 2, pp. 57–64, Sep./Oct. 2015.

[24] T. Hackel, J. D. Wegner, and K. Schindler, "Fast semantic segmentation of 3D point clouds with strongly varying density," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 3 pp. 177–184, Jul. 2016.

[25] 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.

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

[27] 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. Automat.*, Shanghai, China, May 2011, pp. 2609–2616.

[28] S. Ross, D. Munoz, M. Hebert, and J. A. Bagnell, "Learning message-passing inference machines for structured prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Colorado Springs, CO, USA, Jun. 2011, pp. 2737–2744.

[29] F. Tombari, N. Fioraio, T. Cavallari, S. Salti, A. Petrelli, and L. Di Stefano, "Automatic detection of pole-like structures in 3D urban environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2014, pp. 4922–4929.

[30] 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.

[31] D. Wolf, J. Prankl, and M. Vincze, "Fast semantic segmentation of 3D point clouds using a dense crf with learned parameters," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2015, pp. 4867–4873.

[32] H. S. Koppula, A. Anand, T. Joachims, and A. Saxena, "Semantic labelling of 3-D point clouds for indoor scenes," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Granada, Spain, 2011, pp. 244–252.

[33] S.-M. Hu, J.-X. Cai, and Y.-K. Lai, "Semantic labeling and instance segmentation of 3D point clouds using patch context analysis and multiscale processing," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 7, pp. 2485–2498, Jul. 2020.

[34] Y.-P. Xiao, Y.-K. Lai, F.-L. Zhang, C. Li, and L. Gao, "A survey on deep geometry learning: From a representation perspective," *Comput. Vis. Media*, vol. 6, no. 2, pp. 113–133, Jun. 2020.

[35] D. Maturana and S. Scherer, "VoxNet: A 3D convolutional neural network for real-time object recognition," in *Proc. IROS*, Oct. 2015, pp. 922–928.

[36] J. Huang and S. You, "Point cloud labeling using 3D convolutional neural network," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 2670–2675.

[37] Z. Wu *et al.*, "3D ShapeNets: A deep representation for volumetric shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1912–1920.

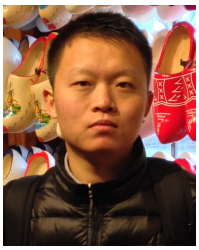
[38] L. Tchapmi, C. Choy, I. Armeni, J. Gwak, and S. Savarese, "SEGCloud: Semantic segmentation of 3D point clouds," in *Proc. Int. Conf. 3D Vis. (3DV)*, Qingdao, China, Oct. 2017, pp. 537–547.

[39] Y. Zhou and O. Tuzel, "VoxelNet: End-to-end learning for point cloud based 3D object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4490–4499.

[40] G. Riegler, A. O. Ulusoy, and A. Geiger, "OctNet: Learning deep 3D representations at high resolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3577–3586.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[41] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," in *Proc. IEEE ICCV*, Dec. 2015, pp. 945–953.
[42] A. Boulch, J. Guerry, B. Le Saux, and N. Audebert, "SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks," *Comput. Graph.*, vol. 71, pp. 189–198, Apr. 2018.
[43] Y. Lu, M. Zhen, and T. Fang, "Multi-view based neural network for semantic segmentation on 3D scenes," *Sci. China Inf. Sci.*, vol. 62, no. 12, pp. 1–3, Dec. 2019.
[44] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 652–660.
[45] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: Convolution on X-transformed points," in *Proc. NIPS*, 2018, pp. 820–830.
[46] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. NIPS*, 2017, pp. 5099–5108.
[47] S. Shi, X. Wang, and H. Li, "PointRCNN: 3D object proposal generation and detection from point cloud," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 770–779.
[48] C. R. Qi, O. Litany, K. He, and L. Guibas, "Deep Hough voting for 3D object detection in point clouds," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9277–9286.
[49] W. Shi and R. Rajkumar, "Point-GNN: Graph neural network for 3D object detection in a point cloud," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1711–1719.
[50] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. Guibas, "KPConv: Flexible and deformable convolution for point clouds," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6411–6420.
[51] K. Mo *et al.*, "PartNet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 909–918.
[52] Q. Hu *et al.*, "RandLA-Net: Efficient semantic segmentation of large-scale point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11108–11117.
[53] Q. Huang, W. Wang, and U. Neumann, "Recurrent slice networks for 3D segmentation of point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2626–2635.
[54] L. Landrieu and M. Simonovsky, "Large-scale point cloud semantic segmentation with superpoint graphs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4558–4567.
[55] W. Wang, R. Yu, Q. Huang, and U. Neumann, "SGPN: Similarity group proposal network for 3D point cloud instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2569–2578.
[56] F. Li, S. O. Elberink, and G. Vosselman, "Pole-like street furniture decomposition in mobile laser scanning data," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 3, no. 3, pp. 193–200, Jun. 2016.
[57] P. Krähenbühl and V. Koltun, "Parameter learning and convergent inference for dense random fields," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 513–521.
[58] F. Li, S. O. Elberink, and G. Vosselman, "Semantic labelling of road furniture in mobile laser scanning data," *ISPRS, Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W7, pp. 247–254, Sep. 2017.
[59] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2002.
[60] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," in *Proc. NIPS*, 2013, pp. 431–439.
[61] C. M. Bishop, *Pattern Recognition and Machine Learning*. NewYork, NY, USA: Springer-Verlag, 2006.
[62] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, May 2003.



Fashuai Li received the Ph.D. degree in photogrammetry and computer vision from the Earth Observation Science Department, Faculty of Geo-Information Science and Earth Observation Science (ITC), University of Twente, The Netherlands, in 2019. He is currently a Research Assistant Professor with the State Key Laboratory of CAD&CG, Zhejiang University. He has published various papers in the leading remote sensing journals and ISPRS events. His research focuses on computer vision, machine learning, and robotics.

Zhize Zhou received the B.E. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2008, the M.Sc. degree from Peking University, Beijing, China, in 2011, and the Ph.D. degree from The University of Sheffield, England, in 2018. He is currently a Research Assistant Professor with the State Key Laboratory of CAD&CG, Zhejiang University. His research interests include solid-state physics, 3D computer vision, and computer graphics.

Jianhua Xiao is currently the Dean of the Wuhan Geomatics Institute. He used to be an Endowed Professor at Wuhan University. His research interests include smart cities, change detection, 3D GIS modeling, and cartography. He is the Chair of the Editorial Board of Urban Geotechnical Investigation and Surveying.

Ruizhi Chen is currently the Director of the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He was an Endowed Chair and a Professor at Texas A&M University-Corpus Christ Campus, USA, and the Head and a Professor of the Department of Navigation and Positioning, Finnish Geodetic Institute, Finland. His research interests include smartphone positioning indoors/outdoors, context awareness, and satellite navigation. He is the General Chair of the IEEE conferences "Ubiquitous Positioning, Indoor Navigation and Location-Based Services," the Editor-in-Chief of *The Journal of Global Positioning Systems*, and an Associate Editor of *Journal of Navigation*.

Matti Lehtomäki is currently pursuing the Ph.D. degree with Aalto University, Espoo, Finland. He is currently a Research Scientist with the Finnish Geospatial Research Institute, Helsinki, Finland. His research interests include computational methods, algorithm development, laser scanning, and computer vision.

Sander Oude Elberink received the Ph.D. degree in 2010. Since September 2009, he has been an Assistant Professor with the Department of Earth Observation Science, ITC, University of Twente. His main research interests are (semi-)automated acquisition of geoinformation, 3D modeling/reconstruction, and information extraction. He received the Young Author's Award for Best Paper at the ISPRS Congress, Beijing, China, in 2008. Since 2012, he has been the Chair of WG III/2, ISPRS.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

George Vosselman received the M.S. degree in geodetic engineering from the Delft University of Technology, The Netherlands, in 1986, and the Ph.D. degree from the Rheinische Friedrich Wilhelms University of Bonn, Germany, in 1991. After a year as a Visiting Scientist at the University of Washington, USA, he was appointed as a Professor of photogrammetry and remote sensing at the Delft University of Technology in 1993. In 2004, he joined ITC, University of Twente, as a Professor of geo-information extraction and remote sensing. The research of his chair group at ITC focuses on the utilisation of advancements in sensor technology for the efficient production of large scale geo-information. He has published over 150 journals and conference papers on photogrammetry and laser scanning. From 2005 to 2012, he was the Editor-in-Chief of the *ISPRS Journal of Photogrammetry and Remote Sensing*.

Juha Hyyppä is currently a Distinguished Research Professor, the Director of the Centre of Excellence in Laser Scanning Research (Centre grant by Academy of Finland), and the Director of the Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, Helsinki, Finland. His research interests include laser scanning systems, their performance and new applications, especially related to mobile and ubiquitous laser scanning systems, including autonomous driving and point cloud processing.

Yuwei Chen received the B.E. and M.Sc. degrees from Zhejiang University, Hangzhou, China, in 1999 and 2002, respectively, and the Ph.D. degree in circuits and systems from the Shanghai Institute of Technical Physics (SITP), Chinese Academy of Sciences, Beijing, China, in 2005. He is currently working with the Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, Masala, Finland, as a Research Manager, where he is leading the Research Group Remote Sensing Electronics, with a focuses on developing new remote sensing systems. He holds ten patents and has authored and coauthored more than 120 scientific papers and book chapters. His research interests include spaceborne LiDAR, hyperspectral LiDAR, radar technology, and seamless navigation.

Antero Kukko is currently a Research Professor with the Centre of Excellence in Laser Scanning Research, Finnish Geospatial Research Institute, Helsinki, Finland, and also an Adjunct Professor with the Department of Built Environment, Aalto University, Espoo, Finland. His research interests include the development and use of mobile laser scanning systems in multiple applications, from impact cratering and fluvial processes to precision forestry, urban mapping, and road conditions and assets.