ISPRS Journal of Photogrammetry and Remote Sensing 154 (2019) 98–113

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Semantic segmentation of road furniture in mobile laser scanning data

Fashuai Li[a,b,c,*], Matti Lehtomäki[b,c], Sander Oude Elberink[a], George Vosselman[a], Antero Kukko[b,c], Eetu Puttonen[b,c], Yuwei Chen[b,c], Juha Hyyppä[b,c]

[a] Faculty of Geo-Information Science and Earth Observation (ITC), Department of Earth Observation Science, University of Twente, Enschede, the Netherlands
[b] Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute, Masala, Finland
[c] Centre of Excellence in Laser Scanning Research, Academy of Finland, Helsinki, Finland

ARTICLE INFO

Keywords:
Pole-like road furniture
Interpretation
Decomposition
Machine learning classifiers
Mobile laser scanning

ABSTRACT

Road furniture recognition has become a prevalent issue in the past few years because of its great importance in smart cities and autonomous driving. Previous research has especially focussed on pole-like road furniture, such as traffic signs and lamp posts. Published methods have mainly classified road furniture as individual objects. However, most road furniture consists of a combination of classes, such as a traffic sign mounted on a street light pole. To tackle this problem, we propose a framework to interpret road furniture at a more detailed level. Instead of being interpreted as single objects, mobile laser scanning data of road furniture is decomposed in elements individually labelled as poles, and objects attached to them, such as, street lights, traffic signs and traffic lights.

In our framework, we first detect road furniture from unorganised mobile laser scanning point clouds. Then detected road furniture is decomposed into poles and attachments (e.g. traffic signs). In the interpretation stage, we extract a set of features to classify the attachments by utilising a knowledge-driven method and four representative types of machine learning classifiers, which are random forest, support vector machine, Gaussian mixture model and naïve Bayes, to explore the optimal method. The designed features are the unary features of attachments and the spatial relations between poles and their attachments. Two experimental test sites in Enschede dataset and Saunalahti dataset were applied, and Saunalahti dataset was collected in two different epochs. In the experimental results, the random forest classifier outperforms the other methods, and the overall accuracy acquired is higher than 80% in Enschede test site and higher than 90% in both Saunalahti epochs. The designed features play an important role in the interpretation of road furniture. The results of two epochs in the same area prove the high reliability of our framework and demonstrate that our method achieves good transferability with an accuracy over 90% through employing the training data of one epoch to test the data in another epoch.

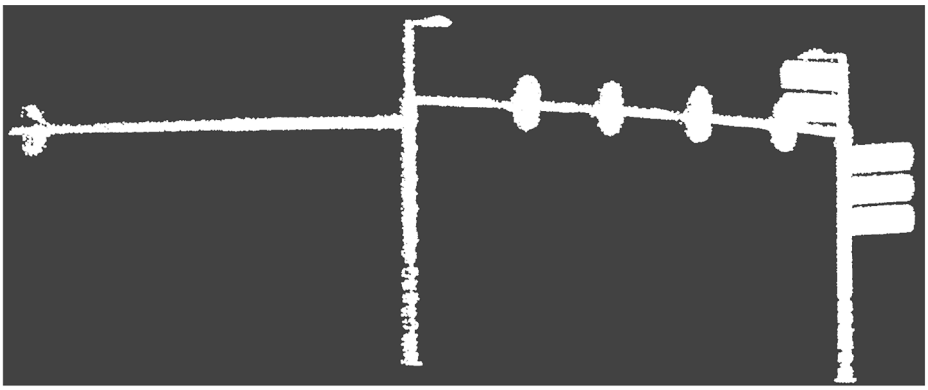
1. Introduction

In recent years, the automatic recognition of road furniture is in great demand because of the pivotal role of road furniture in the road safety analysis. However, road furniture recognition has relied on manual labelling, which is tedious and time-consuming. The emerging of mobile laser scanning provides a solution to this problem. Mobile laser scanning systems have been well developed in recent years, which makes it possible to capture 3D point clouds of urban scenes with higher precision (Ma et al., 2018). Compared to laser scanning systems mounted on other platforms (e.g. terrestrial and airborne laser scanners), mobile laser scanning systems are able to collect 3D data in an urban environment more precisely than airborne laser scanning systems and more rapidly than terrestrial laser scanning systems.

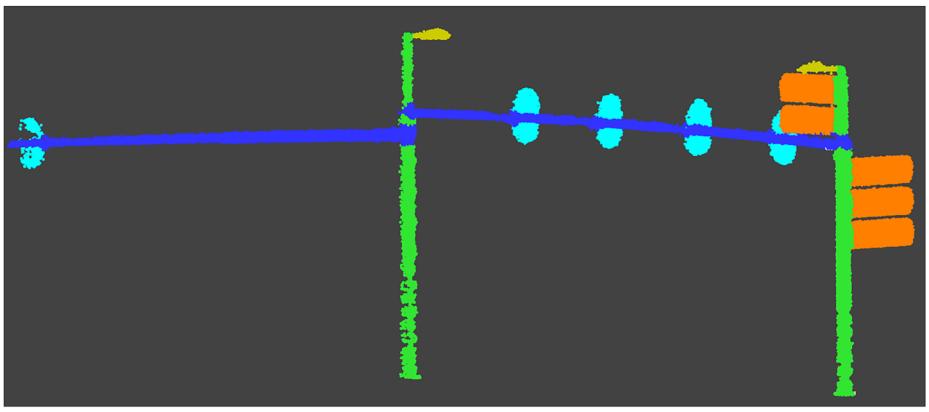
With the support of high-quality mobile laser scanning (MLS) data, road furniture recognition has been well-studied in the past few years. A primary current focus in road furniture inventory is pole-like road furniture identification. Numerous studies have been conducted on pole-like road furniture recognition (Brenner, 2009; Golovinskiy et al., 2009; Lehtomäki et al., 2010; Pu et al., 2011; Xiong et al., 2011; El-Halawany and Lichti, 2013; Bremer et al., 2013; Li and Oude Elberink, 2013; Cabo et al., 2014; Fukano and Masuda, 2015; Huang and You, 2015; Yang et al., 2015; Yu et al., 2015; Li et al., 2017; Wang et al., 2017). The performance of pole-like road furniture classification has improved

* Corresponding author.
E-mail addresses: f.li@utwente.nl (F. Li), matti.lehtomaki@nls.fi (M. Lehtomäki), s.j.oudeelberink@utwente.nl (S. Oude Elberink), george.vosselman@utwente.nl (G. Vosselman), antero.kukko@nls.fi (A. Kukko), eetu.puttonen@nls.fi (E. Puttonen), yuwei.chen@nls.fi (Y. Chen), juha.hyyppa@nls.fi (J. Hyyppä).

https://doi.org/10.1016/j.isprsjprs.2019.06.001
Received 5 November 2018; Received in revised form 31 May 2019; Accepted 3 June 2019



(a)



(b)

Fig. 1. A piece of road furniture with multiple classes. (a) The original point cloud of one piece of road furniture. (b) The interpreted road furniture (Orange: Street signs, Yellow: Street lights, Cyan: Traffic lights, Green: vertical poles, Blue: Horizontal poles). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

significantly in recent years. Most studies focus on the classification of road furniture as single objects by performing a connected component analysis with above-ground points, whereas little attention has been paid to the interpretation of single pieces of road furniture with multiple classes, such as signs and lamps attached to poles. Fig. 1a shows an example where road furniture can only be interpreted as a single object by current frameworks. Fig. 1b shows a detailed interpretation of such road furniture, which is more appropriate and of more practical use. To achieve this goal, we propose a framework to interpret pole-like road furniture. In the first stage, road furniture is detected from MLS data. Then a decomposition stage is carried out to separate detected road furniture into poles and their attachments (e.g. traffic signs and lamps). In the final stage, these attachments are classified into multiple classes by using extracted features and machine learning classifiers. Literally, the point cloud of an attachment is a segment. Our contribution is to fully enable the potential of a decomposed point cloud to understand which classes can be assigned to which attachment. Because of the limited number of pieces of road furniture, the number of training samples is quite small in our research. Therefore, we chose not to use deep neural networks. The performance of different machine learning classifiers is tested on two test sites. We also compare machine learning classifiers to our previously proposed knowledge-driven method (Li et al., 2017). The performance of road furniture interpretation has been significantly improved compared to the knowledge-driven method.

2. Related work

In this section, we divide studies on road furniture recognition into knowledge-driven and machine learning-based methods. We also treat a branch of machine learning, deep learning as a separate group.

Knowledge-driven methods represent early attempts to recognise objects alongside roads from MLS data. This type of methods utilises generic rules to classify target objects. Techniques used for the recognition of structures in point clouds have been reviewed by Vosselman et al. (2004), which includes the recognition of smooth surfaces, planar surfaces and parameterised shapes. Brenner (2009) use a local scatter matrix and a 3D cylinder stack to detect pole-like road furniture from MLS data. In this method, the generic rule of pole-like road furniture detection is to analyse the distribution of points in both the inner and outer region of a fitted cylinder. Utilising similar rules, Lehtomäki et al. (2010) develop a framework to detect vertical-pole objects by using a scan line segmentation strategy. A voxel-based algorithm is proposed by Cabo et al. (2014) to detect pole-like street furniture objects from MLS data by using a similar cylinder masking. Bremer et al. (2013) employ two radius neighbourhoods to extract linear, planar and volumetric features. Rules are defined to classify objects into three categories in this piece of work. However, it is difficult to extract slanted poles and thick poles like tree trunks for this method. A percentile-based method is proposed by Pu et al. (2011) to identify pole-like road furniture from MLS data for road inventory studies. Traffic signs are differentiated by using shape information in this method. Li and Oude Elberink (2013) optimise the work of Pu et al. (2011) by the additional use of reflectance information. Because of the addition of reflectance information, the detection rate of traffic signs is significantly improved. Similar to Pu et al. (2011), shape information is employed by Riveiro et al. (2016) to recognise traffic signs. However, these three methods have difficulties with the detection of road furniture with complex structures, such as road furniture connected with other objects. A 2D point density feature is exploited by El-Halawany and Lichti (2013) to identify road poles from mobile terrestrial laser scanning data. Oude Elberink and Kemboi (2014) recognise objects by using the slicing features and template matching in MLS data. Yang et al. (2015) introduce a supervoxel-based method to classify urban objects from MLS data. Knowledge-driven methods rely on generic rules

which are obtained from the inference of empirical observation. It is not appropriate for this type of method to perform classification tasks with many target classes. This is because the number of generic rules in combination with feature constraints is somewhat limited. This type of methods does not require much training data.

Machine learning based methods have been widely applied for point cloud processing in recent years. Golovinskiy et al. (2009) propose a shape-based approach to recognise 3D point clouds in urban environments by using the min-cut algorithm. Munoz et al. (2009) adopt Max-Margin Markov Networks (M3N) to classify urban scene objects into five categories. Ross et al. (2011) compare five message-passing approaches for predicting labels of point clouds. Their proposed message-passing algorithm shows the scalability and efficacy with 3D point cloud classification. Based on their work, Xiong et al. (2011) propose a sequenced predictor to do 3-D scene analysis. However, the precision of pole and tree trunk recognition using M3N is low compared with the identification of other categories. Yu et al. (2015) present a method to separate urban road objects and recognise street lights from MLS point clouds using the normalised cut algorithm. Velizhev et al. (2012) present an implicit shape models (ISM) based method to localise and recognise cars and light poles automatically. Spin image (SI) descriptor is employed as the keypoint representation for recognition. Yokoyama et al. (2013) propose a method to detect and classify pole-like road furniture from MLS data. Both shape features of pole-like objects and the distribution of surrounding pole-like objects are used in this method. The support vector machine classifier in combination with a set of handcrafted features is employed to classify point clouds of urban scenes (Yang and Dong, 2013; Huang and You, 2015; Soilán et al., 2016; Lehtomäki et al., 2016). Random forest is adopted with manually drafted features to identify objects from MLS data (Fukano and Masuda, 2015; Hackel et al., 2016). Weinmann et al. (2015) compare different machine learning classifiers for the classification of urban environment objects by using features extracted with an optimal neighbourhood size. Tombari et al. (2014) combine local descriptors and global descriptors to recognise pole-like road infrastructures automatically. Support vector machine and Markov random fields (MRF) are adopted at a local level and global level, respectively. The classification of pole-like road furniture both in point-level and in object-level has been improved significantly by machine learning classifiers in combination with manually designed features because of the acquisition of higher quality data.

Deep learning has been studied and applied to object recognition in 3D point clouds. Based on the representation of point clouds, current deep learning techniques on point cloud classification are categorised as three types: convolutional neural networks (CNNs) operating on voxels, point clouds and multi-view images. Voxel-based methods regularise point clouds by using voxels. Maturana and Scherer (2015) develop VoxNet, convolutional neural network (CNN) to train and recognise objects in real-time 3D point clouds. VoxNet first transforms original point clouds to 3D voxel grids. Then a 3D CNN is constructed to consume these regular voxels as an input to train a model and predict the labels of point clouds. Huang and You (2016) applied VoxNet to classify objects in 3D point clouds of urban scenes. Riegler et al. (2017) conduct the classification of sparse 3D point clouds by building a 3D CNN on the basis of the octree representation. In contrast to voxel-based methods, methods that use CNNs operating on point clouds use transformed voxels or supervoxels instead of directly using regular voxels to feed neural networks. A Bag of Words (BoW) and Deep Boltzmann Machine (DBM) method is applied to detect and recognise traffic signs in MLS data by Yu et al. (2016). PointNet is proposed by Qi et al. (2017) to semantically segment scene and classify objects in 3D point cloud by respecting the permutation invariance of 3D points. PointNet directly utilises point clouds as an input. Then a symmetry function for the unordered input is proposed to aggregate the formation for each point and make the model invariant to input permutation. These point features are aggregated by max pooling. In the end, the output is the classification score for each class. Landrieu and Simonovsky (2017) embed PointNet with gated recurrent unit to construct graphs, which segment large-scale point cloud semantically. Similar to the PointNet, Li et al. (2018c) exploit a 3D CNN to segment a point cloud semantically by learning a transformation. Su et al. (2018) present a network structure to interpret objects in 3D point clouds by representing a collection of points as a sparse set of samples in a high-dimensional lattice. Different from the other two types of method, multi-view image-based methods project point clouds to multiple planes to generate image features for every point. In this way, deep learning techniques, which are employed for images processing, can be applied to process point clouds. Su et al. (2015) project point clouds onto different views to generate 2D rendered images. Then they concatenate different rendered images to train a CNN to recognise objects. Boulch et al. (2017) further develop this type of network to process 3D point clouds in the large scale urban scene. Compared to knowledge-driven methods and traditional machine learning classifiers, deep learning methods are more flexible. There is no need for this type of methods to manually design generic rules and high-level features. Deep learning has been applied to 3D point clouds processing and advanced this field significantly. However, it requires numerous training data with good quality.

Up to now, these methods mainly focus on the classification of entire pieces of road furniture. We have proposed a knowledge-driven method to interpret decomposed pole-like road furniture in MLS data (Li et al., 2017). Otherwise, the detailed interpretation of road furniture has not yet been addressed.

3. Methodology

In this section, we introduce a fully automatic framework to interpret pole-like road furniture based on its functionalities. Original MLS point clouds are partitioned into blocks based on trajectory information in a pre-processing step. At the first stage, pole-like road furniture is detected from these blocks. Detected pole-like road furniture is decomposed into poles and their attachments then. In the end, these separated attachments are interpreted by their functionalities. The workflow of our framework is as shown in Fig. 2.

3.1. Data partitioning

When we process a considerable amount of collected mobile laser scanning data in one go, there are computation and memory problems. To cope with these problems, we split the original point cloud into blocks along the trajectory line as described in the piece of work of Pu et al. (2011). We split a trajectory line into multiple blocks which are oriented along the road directions obtained from trajectory data. A block is constructed by manually specify the length of the trajectory line and the width. This schematic is valid for both straight and curve roads. One block of MLS data is as shown in Fig. 3.

3.2. Pole-like road furniture detection

After the original MLS point cloud is split into blocks, we use the local height variance in combination with the height of the corresponding trajectory points to detect ground points for every block. Normally, points of ground are with low height variance and below the elevation of corresponding trajectory points. The description of this method can be found in our previous work (Li et al., 2018a). This is a simple method. The parameters used are the local height variance and the height of the corresponding trajectory points. The first parameter is defined to be 0.15 m, and the second parameter can be directly obtained from scanning files.

Then the ground points are filtered out, and the above-ground points are obtained. A connected component analysis is subsequently carried out to cluster above-ground points into individual above-ground

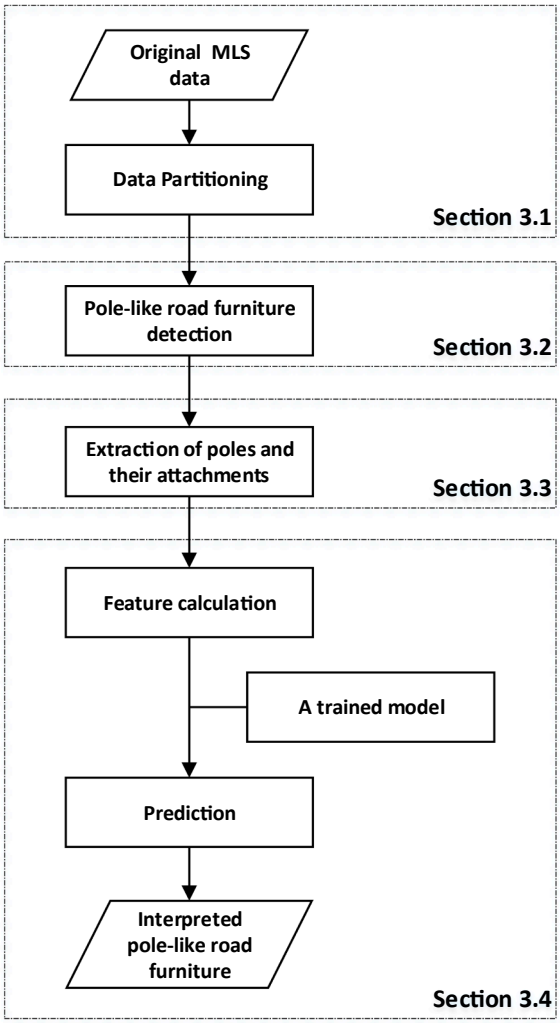


Fig 2. The workflow of pole-like road furniture interpretation.

objects. We perform a rough classification to recognise building, trees and pole-like road furniture then. In this phase, buildings are identified by checking if the clustered objects contain large and vertical planes. To detect planes, we first use the Hough transform to determine the seed points. Then a region growing method is employed to finalise the detection of a plane. The associated parameters are the height, width and the area of a detected plane, and the orientation of the normal of the plane. After buildings are detected and removed, we use the percentage of points with the first returning pulse in a cluster to detect trees. A large number of points not from the first returning pulse suggests that the above-ground cluster is a tree. Edge points have multiple returns. but they also have first return. A part of points of trees are with second return or third return, and without the first return. That's why here we use the percentage of points not from the first return. Trees are then identified and removed from above-ground objects. At the end of this stage, we extract pole-like road furniture by cutting the remainder above-ground objects into slices and performing 2D connected components analysis with the centre points of these slices. The details of this pole-like road furniture detection can be found in (Li et al., 2018b). Fig. 4 shows detected road furniture in one block of MLS data.

3.3. Pole-like road furniture decomposition

In this stage, we aim to decompose detected pole-like road furniture into poles and attachments by using a two-step method. The first step is to extract poles from every piece of pole-like road furniture, and we separate attachments which are connected to poles in the second step. In this paper, attachments are any parts of pole-like road furniture which are not extracted as horizontal or vertical poles.

Pole-like road furniture is first classified into three types. And they are vertical poles connected with many attachments, vertical poles connected with few attachments and vertical poles connected with horizontal poles and attachments. We differentiate the first and the second type of pole-like road furniture by comparing the median width and the lowest quartile width of cut slices. Based on the type of road furniture, a respective method is automatically selected to extract poles from the objects. Then the pole extraction method is performed to obtain the points which belong to poles. After poles are extracted and removed, a connected component analysis is carried out to cluster points to different attachments. In order to achieve a better decomposition, a set of rules is designed to merge and split attachments. A detailed description of pole-like road furniture decomposition can be found in (Li et al., 2018a). Decomposed road furniture in one block of MLS data is as illustrated in Fig. 4.

3.4. Pole-like road furniture interpretation

3.4.1. Feature notation

Different from pointwise features, features are extracted from attachments in relation to poles, which are regarded as units in this paper. Therefore, these utilised features are segment-wise. In addition to unary features, contextual features are extracted as well to feed machine learning classifiers for training and predicting. In this paper, unary

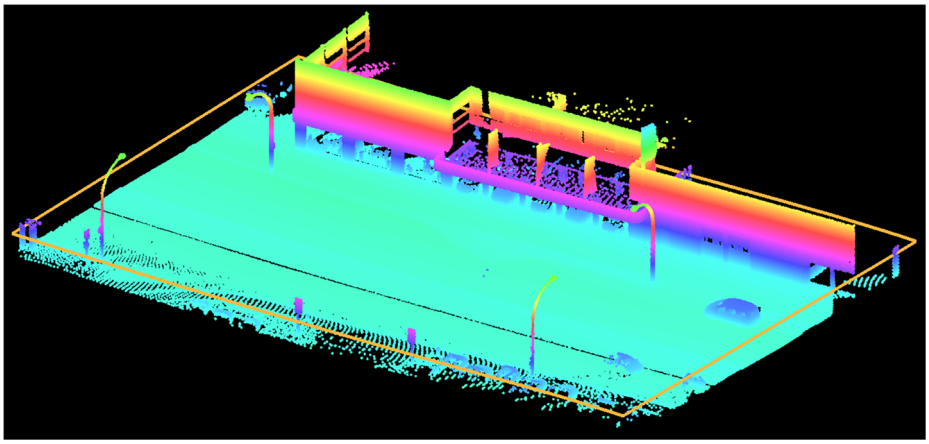


Fig. 3. One block of MLS data (points are coloured by their elevation). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

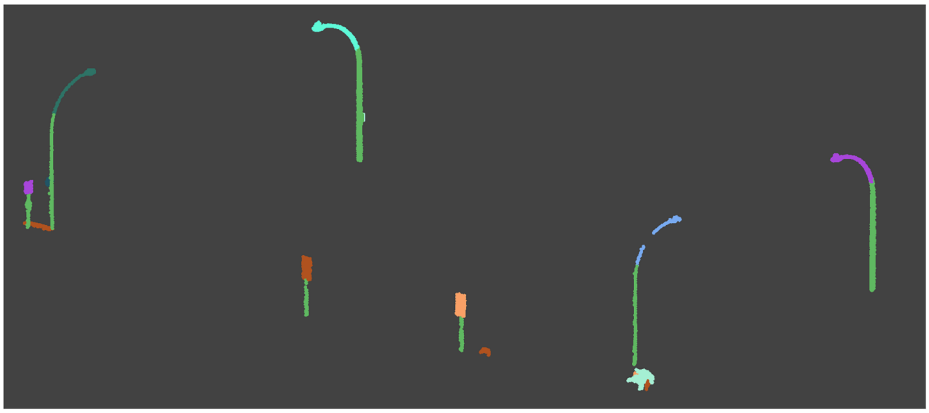


Fig. 4. Detected and decomposed pole-like road furniture in one block of MLS data (green points denote poles, and points with other colours imply attachments). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

features are features which are directly obtained from attachments. Three types of unary features are extracted. They are size $\{S\}$, eigenvalue-based features $\{E\}$ and radiometric features $\{R\}$ of attachments.

To extract size features $\{S\}$, we construct the bounding box for every attachment as shown in Fig. 5. A six dimensional feature $\{S\}$ is calculated for the bounding box. It includes the horizontal length S_L, horizontal width S_W, height S_H, the maximum range in the horizontal plane S_D, volume $S_V = S_L \cdot S_W \cdot S_H$ and ratio feature S_{D-H}. S_D is the largest distance between two points of an attachment in the horizontal plane. Ratio feature S_{D-H} is the ratio of the 2D distance S_D to the height of the bounding box S_H. When we compute the largest 2D distance S_D of two points in the point cloud of an attachment, the complexity of the calculation of this feature is $O(n^2)$.

Eigenvalue-based features $\{E\}$ are calculated by eigenvalues $\{\lambda_1 > \lambda_2 > \lambda_3 \geq 0\}$ of the covariance matrix of point coordinates belonging to every attachment. They are linearity E_L, planarity E_P, linear

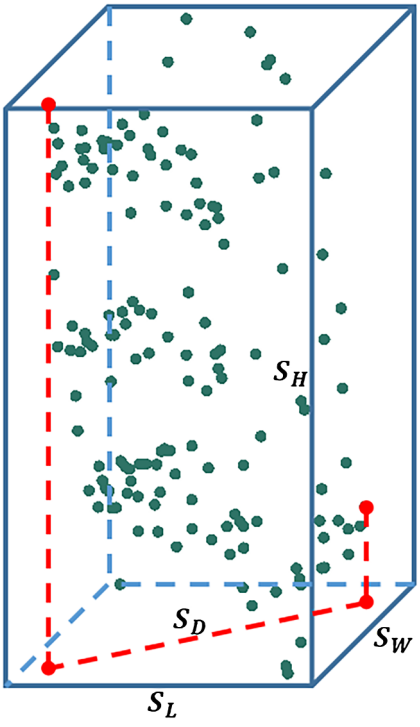


Fig. 5. The size features of a traffic light (the blue box is the bounding box of the point cloud of the traffic light). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

planarity E_{LP}, scattering E_S, omnivariance E_O, anisotropy E_A, eigenentropy E_E, and change of curvature E_C. E_L describes the linearity of attachment. E_P denotes how planar an attachment is. E_{LP} is designed for specifying elongated street signs. E_S depicts how scattered the distribution of points of an attachment is. E_O suggests how spherical an attachment is. E_A implies the distribution of points of an attachment in different directions. The entropy of the distribution of points in three directions is indicated by E_E. E_C denotes the change of curvature of an attachment. The calculation of these features is as Eq. (1) to Eq. (9). Compared to the features used by Weinmann et al. (2015), we add E_{LP}.

$$e_i = \frac{\lambda_i}{\sum_i \lambda_i},\ i \in \{1, 2, 3\} \tag{1}$$

$$E_L = \frac{e_1 - e_2}{e_1} \tag{2}$$

$$E_P = \frac{e_2 - e_3}{e_1} \tag{3}$$

$$E_{LP} = \frac{e_2 - e_3}{e_2} \tag{4}$$

$$E_S = \frac{e_3}{e_1} \tag{5}$$

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

$$E_A = \frac{e_1 - e_3}{e_1} \tag{7}$$

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

$$E_C = e_3 \tag{9}$$

Radiometric features $\{R\} = \{R_{MD}, R_{MA}\}$ are obtained directly from the collected radiometric attributes of the point cloud. R_{MD} and R_{MA} are the median and mean value of reflectance of points in every attachment.

In addition, there are two types of contextual features, which describe the spatial relations between attachment and their attached poles. Unlike the extraction of unary features, contextual features are calculated from attachments and poles. They are angle features $\{A\}$ and relative height features $\{H\}$. The detailed explanation of these features is as follows.

Angle features $\{A\}$ consist of the angle A_{PA} between the principal direction of poles $\vec{V}_P$and the normal direction of their attachments $\vec{V}_A$, and the angle A_{VA} between vertical direction $\vec{V}_V$ and the normal direction of attachments. They are defined in the following two equations:

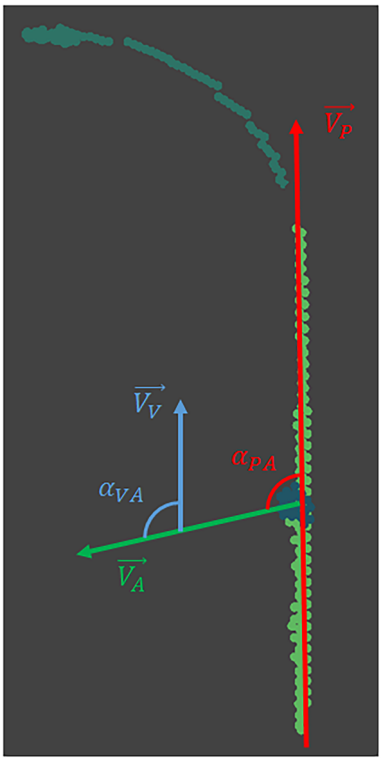


Fig. 6. Angle features (see text for the explanation).

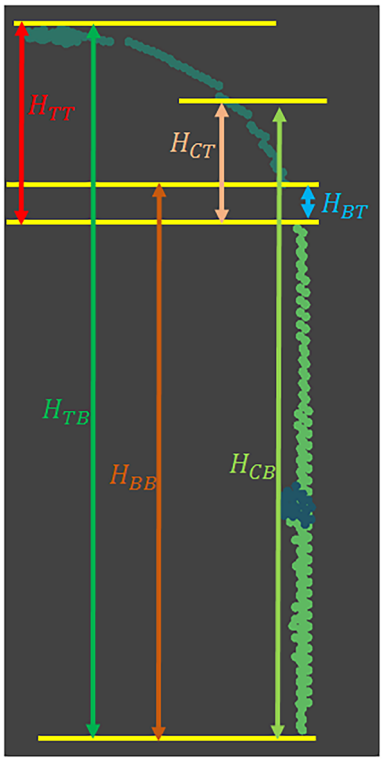


Fig. 7. Height features between a pole and an attachment. Green points denote pole, and the street light and the traffic sign are represented by dark cyan points and steel blue points, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$A_{PA} = \overrightarrow{V_P} \bullet \overrightarrow{V_A} = \cos\alpha_{PA} \tag{10}$$

$$A_{VA} = \overrightarrow{V_V} \bullet \overrightarrow{V_A} = \cos\alpha_{VA} \tag{11}$$

One instance of angle features of a traffic sign is as illustrated in Fig. 6.

Height features $\{H\}$ are the differences of height between the position of attachments and the position of poles. They include H_{CB}, H_{CT}, H_{BB}, H_{BT}, H_{TB}, H_{TT}. The first sub-index is the index of the position of an attachment, and the second sub-index is the index of the position of its connected pole. The sub-index C denotes the centre position, B the lowest position, and T the highest position. For instance, H_{CB} is the height difference between the centre point of the attachment and the lowest point of its connected pole. The illustration of height features is as shown in Fig. 7. The height features are not only applicable for the demonstrated type of pole. Except for curved poles, this type of feature works with any type of straight poles.

3.4.2. Machine learning classifiers

To classify the separated attachments, we employ four different types of machine learning classifier: support vector machine (SVM), random forest (RF), Gaussian mixture model (GMM) and naïve Bayes (NB). First, the classifiers are trained using training data with features extracted in Section 3.4.1. Then class labels are predicted for new unseen data using the trained classifiers. In this paper, we apply the C-SVM classifier (Chang and Lin, 2011; Bishop, 2006) and use the radial basis function (RBF) kernel, suggested in Hsu et al. (2003). The LIBSVM package (Chang and Lin, 2011) is used to train the classifiers and to perform predictions. As suggested in Hsu et al. (2003) and Hsu and Lin (2002), all features are scaled between 0 and 1, and the one-versus-one approach is used for multi-class classification.

C-SVM that uses RBF kernel contains two hyperparameters γ and C that control the complexity of the classifier. To find optimal values for γ and C, cross-validation (CV) (Kohavi, 1995) in combination with grid search is used. The γ-C pair with the highest CV accuracy was selected as an optimal model. In the end, the optimal model obtained from cross validation is used to evaluate the performance of the classification model on an independent test set.

To improve the classification performance and to analyse the goodness of different extracted features, also feature selection is performed. We use backward elimination (Guyon and Elisseeff, 2003), as it is able to take into account joint effects of various features on the classification accuracy.

The random forest classifier as an ensemble learning method takes advantage of the combination of multiple weak learners. It has been widely employed in classification applications (Criminisi et al., 2012). Bootstrap aggregating in combination with information gain is utilised to construct a group of trees called random forest (Breiman, 2001). Two hyperparameters, the number of trees N and the depth of trees D, exist in the random forest classifier. The optimal combination of these two hyperparameters is selected using a 2D grid search. In order to analyse the utility of our designed features, the importance of every feature is also computed by the Gini index, which is described in Louppe et al. (2013).

The Gaussian mixture model is often used in combination with k-means clustering, which gives initial clustering information. Instead of using the k-means clustering, our training data is manually labelled. In GMM, the distribution of an attachment belonging to a class follows a multivariate Gaussian distribution. We used the Expectation Maximization (EM) to estimate all the mixture parameters. In the end, the label of an attachment is given to the class with the highest probability.

As a method based on Bayes' theorem, the naïve Bayes classifier assumes every pair of features is conditionally independent given the class (Murphy, 2012). It is also assumed that every feature given their labels follows a Gaussian distribution. Based on Bayes' theorem and Maximum A Posteriori (MAP), the probability of the label of an attachment is estimated. In the end, every attachment is given the label with the highest probability.

Fig. 8. The Optech LYNX mobile mapping system.

3.4.3. The extraction of non-changed attachments in two epochs

In order to test the robustness of our framework, we test the difference of predictions of non-changed attachments in two epochs. In this paper, non-changed attachments are attachments appearing in both epochs of the Saunalahti dataset in the same position. To extract the non-changed attachments, we first detect non-changed poles and obtain their attachments. The non-changed poles are extracted by comparing the 2D distance of centre points of poles from two epochs. If the distance is less than a predefined threshold, they are identified as non-changed poles. The connected attachments of the non-changed poles are obtained afterwards. Then we compare the centre position of bounding boxes of the obtained attachments from two different epochs. If the distance of centre positions is less than a threshold, they are detected as the non-changed attachments. Otherwise, they are identified as changed attachments.

4. Study area and experimental results

4.1. Test sites

To conduct the experiment, we choose two test sites to evaluate the performance of our framework. In both datasets, there are many types of road furniture scanned.

The first MLS dataset was collected in Enschede, a medium-sized city located in the eastern side of Netherlands. In December of 2008, it was collected by TopScan GmbH, which utilised the Optech LYNX mobile mapping system. In this system, there are two laser scanners amounted at the backside of a driving vehicle. The directions of the rotating plane of these two laser beams are perpendicular to each other and form a 45° angle to the driving direction. An example of our adopted system is as shown in Fig. 8. The frequency of both laser scanners was 100 kHz. The maximum driving speed of this platform was 50 km/h. In this dataset pulse count and reflectance were also collected.

The Enschede MLS trajectory is about 5.5 km long along which more than one hundred million points were recorded. The point density ranges from 35 to 350 points per square metre. There are six scanning strips used as training data. In this paper, one scanning strip is the MLS data corresponding to a continuously scanned street trajectory. There are 210 pieces of road furniture in the training dataset. Another four scanning strips, which contains 260 pieces of road furniture, are taken as a testing area. The size of every block is set to be 50 m long and 40 m wide. The strip overview of this dataset is as shown in Fig. 9.

The second test site is located in the Saunalahti district of Espoo, the second largest city in Finland. It was collected in two different epochs. The first epoch was collected in December 2016, and the dataset is named Saunalahti winter. The second dataset, named Saunalahti spring, was collected in May 2017. Both datasets were collected using an FGI proprietary MLS system for road and urban environment mapping (Fig. 10). The system was equipped with a Riegl VUX-1HA laser scanner for the collection of 3D point clouds. The scanner was set to

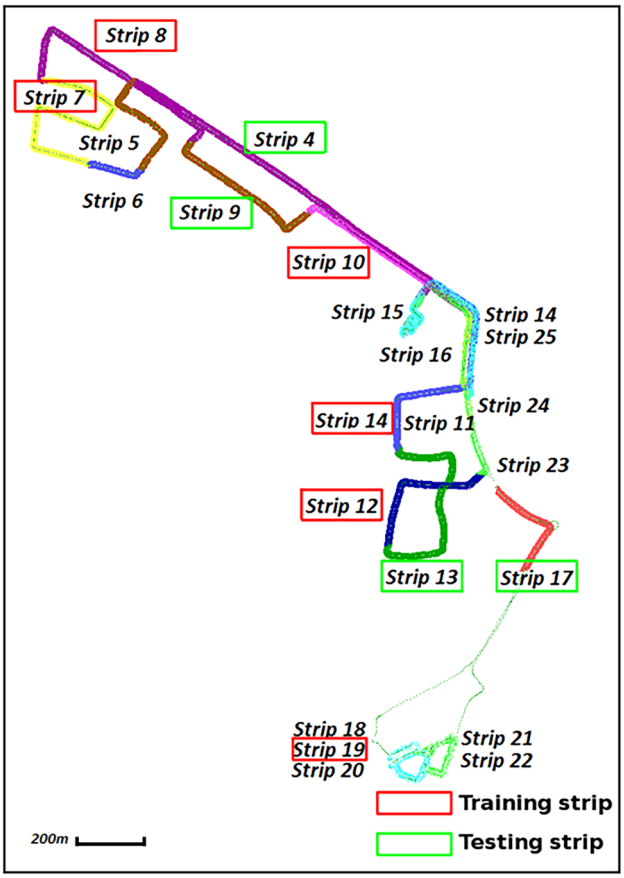


Fig. 9. Enschede dataset. The strips without red or green outlines are not utilised in the experiment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

scan at 250 Hz frequency for cross-track profiles and transmit 1,017,000 pulses per second. The scanner is capable of detecting up to seven echoes per pulse on-line, depending on the object characteristics. The setup resulted in 4068 transmitted pulses per 360-degree profile giving thus 1.5 mrad angular resolution. The area was captured mainly from streets at normal urban speed limits 20–40 km/h. At the approximate sensor elevation of about 2.8 m from the ground, the point spacing right below the sensor was about 4 mm; at 50 m distance from the sensor the point spacing on a perpendicular plane was 77 mm, and the respective footprint size is 15 mm (0.3 mrad divergence).

The Saunalahti winter dataset covers about 5.0 km of road, and the spring dataset covers about 4.0 km of road, consisting of more than one hundred million points. The distance of common area, of which the data was collected in both the winter of 2016 and the spring of 2017, is

Fig. 10. The FGI proprietary mobile mapping system.

about 2.5 km. In the winter dataset, six scanning strips are chosen as training data, and seven strips as testing data. The training datasets and testing datasets contain 111 and 187 pieces of road furniture, respectively. In the Saunalahti spring dataset, four strips are used as training data, which comprises 100 pieces of road furniture. Another six strips are chosen as testing data, in which there are 241 pieces of road furniture. In the data partitioning stage, the size of every block is set to be 50 m long and 40 m wide. The strip overview of this dataset is shown in Fig. 11.

4.2. Results

We evaluate the results of pole-like road furniture interpretation with four aspects. We first compare the performance of the knowledge-driven method and four machine learning classifiers. Then the confusion matrix is given and analysed with the method, which achieves the best performance. The reliability of our framework is evaluated by the difference in the interpretation of corresponding road furniture in two epochs afterwards. In the end, we quantitatively assess the transferability of our framework by swapping trained models.

To produce the ground truth data for the evaluation of interpretation, we only use detected pole-like road furniture. The tag of every attachment is manually labelled. We treat our proposed framework as an automatic procedure. Therefore, undetected pole-like road furniture is not included in the ground truth data. We compare the performance of five different methods for the classification of attachments of pole-like road furniture in the three datasets mentioned above. The first one is the knowledge-driven method (Li et al., 2017). In this method, a set of rules is defined based on the generic mounting patterns of road furniture. For instance, street lights are often mounted on the top of a pole and above a certain minimum height. The details of this method

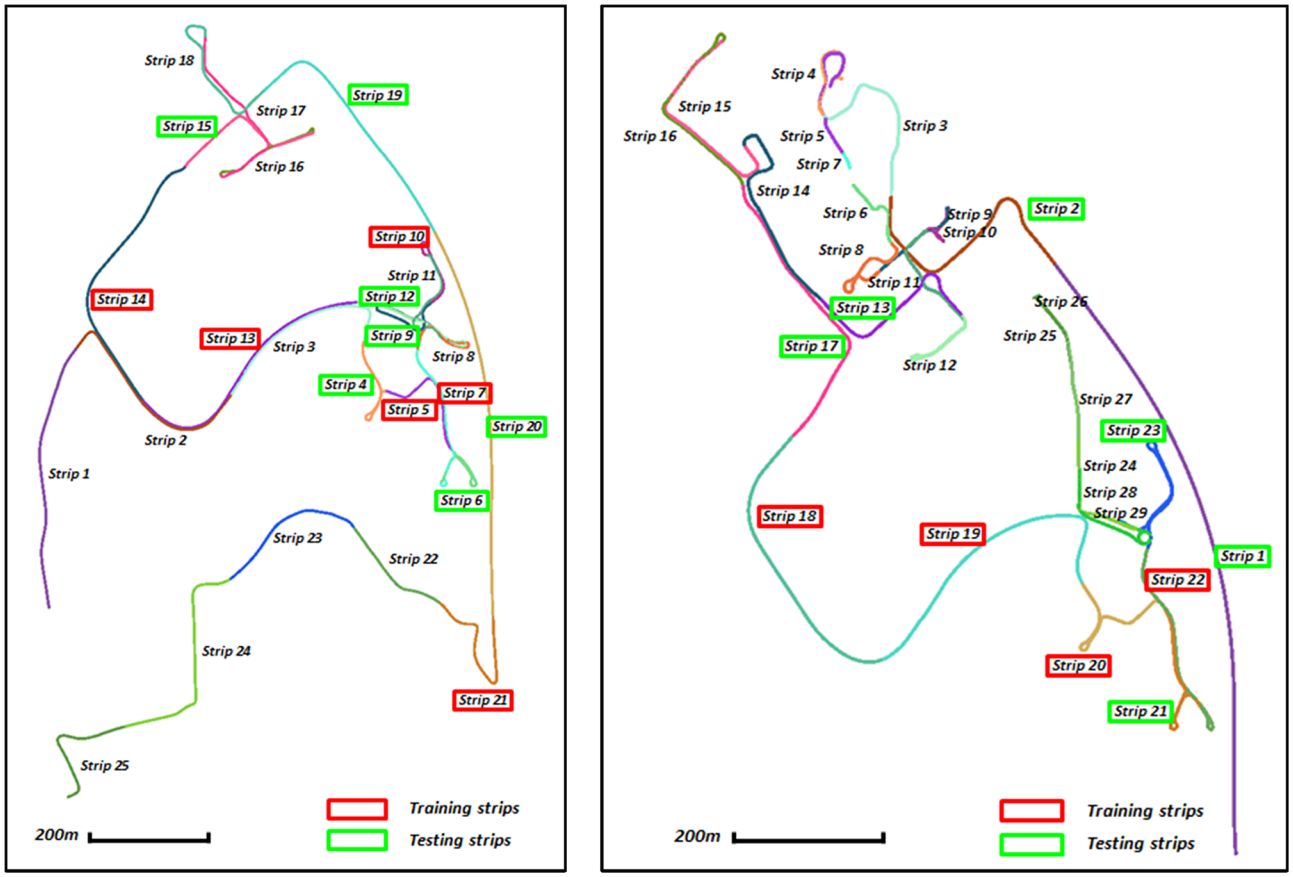


Fig. 11. The strip view of Saunalahti winter dataset (left) and Saunalahti spring dataset (right). The strips without red or green outlines are not utilised in the experiment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
The setting of hyperparameters of machine learning classifiers.

Test sites	SVM		Random forest	
	γ	C	D	N
Enschede	0.05	8	10	20
Saunalahti winter	0.05	5	6	33
Saunalahti spring	0.02	5	6	40

can be found in Li et al. (2017). The other four methods are machine learning classifiers, SVM, random forest, GMM and naïve Bayes introduced in Section 3.4.2. The setting of hyperparameters of machine learning classifiers is as illustrated in Table 1. The quantitative evaluation of the performance is as shown in Table 2. The table indicates that random forest outperforms the other four methods. The precision of correctly interpreted attachments is higher than 80% in Enschede dataset and higher than 90% in both Saunalahti spring and winter datasets. The precision utilised in this paper is the number of correctly recognised attachments divided by the number of all attachments. The ratio of complex pole-like road furniture is almost 50.0% in both Enschede and Saunalahti test sites. Four of the five classification methods show a higher precision of pole-like road furniture interpretation in Saunalahti test site than in Enschede test site. One reason is that there is more complex pole-like road furniture in Enschede test site. The second reason is that there are more types of attachments in Enschede test site. The performance of GMM is the worst. This is because the distribution of feature values (e.g. the size feature of different types of street lights) is not Gaussian. The performance of the knowledge-driven method in Saunalahti test site is worse than the performance in Enschede test site, mainly because the rules have been designed based on the Enschede dataset. The difference of performance of SVM and random forest is analysed in the next section (Section 4.3).

A screenshot of interpreted road furniture (trained and predicted by the random forest model) in these three test sites is as shown in Fig. 12. In this figure, vertical poles and horizontal poles are in green colour, and street lights are coloured to be yellow. Magenta attachments are traffic signs, and orange attachments are street signs. Cyan attachments represent traffic lights, and maroon attachments represent connected traffic signs (Fig. 13). Other signs and other attachments are colourised to be brown and purple, respectively. White attachments are other objects. Fig. 12 indicates that road furniture is well interpreted in both Saunalahti winter and spring datasets.

In addition, we also give the confusion matrices of the results on the three test sites. The confusion matrices for the random forest classifier are as shown in Tables 3, 4 and 5. From the confusion matrix, it can be seen that the accuracy of recognition of street light is the highest. The recognition rate of street signs and connected traffic signs is much lower.

In order to analyse the reliability of our framework, we apply our framework on two different epochs of the Saunalahti dataset and check the difference of interpretation on the non-changed attachments. The detection of non-changed attachments in both epochs of Saunalahti test site is explained in Section 3.4.3. The threshold of the distance between centre points of a pair of corresponding attachments in two epochs is set to be 0.8 m. There are 68 non-changed attachments, among which 60

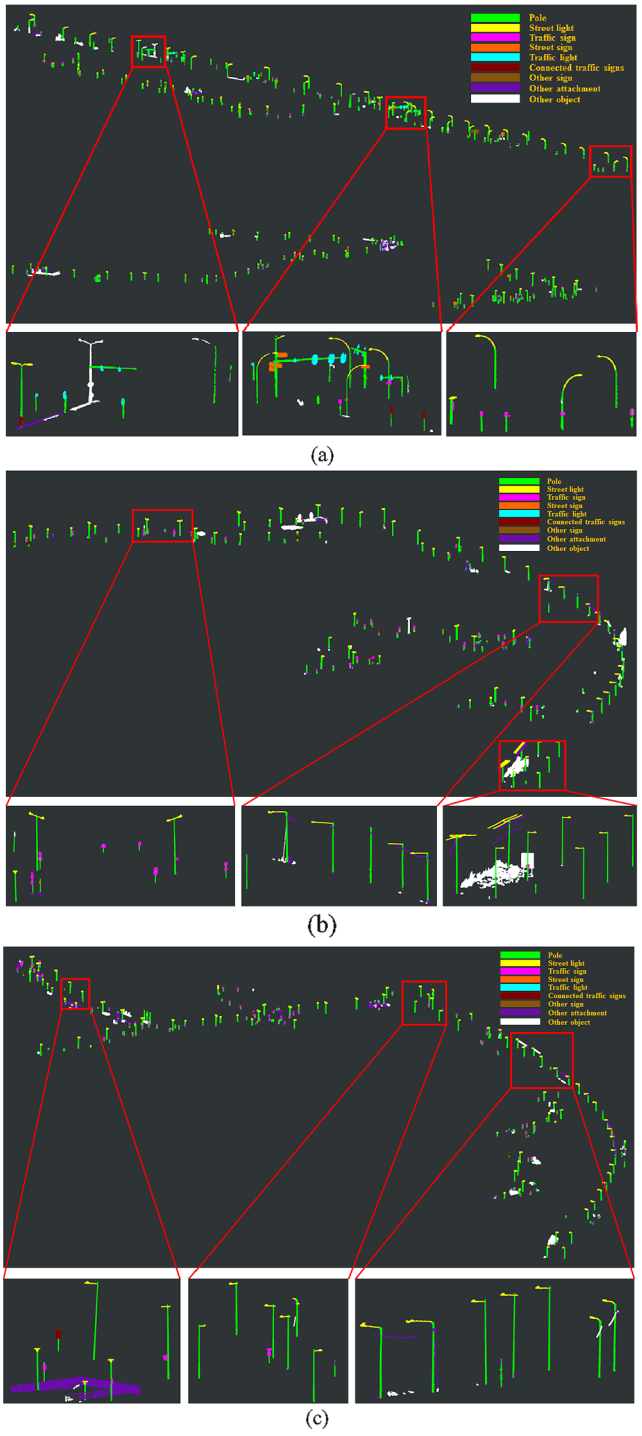


Fig. 12. The visualisation of interpreted road furniture in (a) Enschede dataset, (b) Saunalahti winter dataset and (c) Saunalahti spring dataset.

Table 2
The performance of road furniture interpretation of five different methods.

Test sites	Knowledge-driven	SVM	Random forest	GMM	Naïve Bayes
Enschede	64.4%	72.2%	81.0%	29.0%	59.1%
Saunalahti winter	51.5%	79.1%	92.3%	35.4%	86.8%
Saunalahti spring	59.5%	78.2%	94.1%	40.7%	86.8%

Fig. 13. Connected traffic signs.

are given the same label. The similarity rate of the prediction of the same attachments is 88.2%. Therefore, there is high robustness with our model. The interpretation of non-changed road furniture is as shown in Fig. 14a,b. The 8 objects are predicted with different labels because of incorrect decomposition (as illustrated in Fig. 15a) and stray points (as illustrated in Fig. 15b). In this paper, stray points are noisy points caused by the scanning process. In Fig. 15a, from left to right, the two figures on the left side are the decomposition and the interpretation of road furniture in the Saunalahti winter dataset, and the two figures on the right side are the decomposition and interpretation of road furniture in the Saunalahti spring dataset. The traffic signs in the Saunalahti winter dataset (the magenta attachments in the second left figure of Fig. 15a) are separated as shown in the first left figure of Fig. 15a. Because of the incorrect decomposition in the Saunalahti spring dataset, the upper two attachments are connected with each other (the sea green attachment in the third left figure of Fig. 15a), which leads to the wrong prediction of these two attachments. They are misclassified as a connected traffic sign (the maroon attachment in the fourth left figure of Fig. 15a). In Fig. 15b, the connected traffic sign (the left figure) is correctly predicted in the Sauanalahti winter dataset. There are points resulting from scanning artefacts surrounding the traffic sign in the right figure of Fig. 15b. The calculation of size features and eigenvalue-based features is not reliable, which triggers the wrong prediction.

We also evaluate the transferability of our framework. Transferability in this paper refers to two different situations. First, transferability refers to the capability of applying a model, which is trained with a data that is collected in one epoch to the predict the labels of a dataset collected in another epoch, both epochs in the same area. Second, transferability refers to the capability of applying the trained model in a new area. We evaluate the transferability by testing the performance of the trained model of the Saunalahti winter dataset on the Saunalahti spring dataset, and vice versa. As explained above in Section 4.1, we collected the data in Saunalahti test area in both winter and spring. There are also differences caused by the construction work, and therefore the two datasets are not completely the same. Therefore, they are used for model transferability analysis. The transferability of the random forest model is as shown in Table 6. When the trained model of winter dataset is applied to predict the spring dataset, the accuracy is 92.5%. When the trained model of spring dataset is used to predict the winter dataset, the accuracy is 91.2%. It demonstrates that the transferability of our random forest model is high when two datasets share the same types of road furniture. Aside from the analysis of transferability of different epochs in the same area, we also test the transferability of trained models between different test sites. We use the trained model of Enschede test site to predict the label of attachments in the Saunalahti test site. The overall accuracy of prediction is 65.8% and 69.7% in Saunalahti winter and spring dataset, respectively. The recognition rate of street light is still higher than 87.0%. In Enschede test site, there are seven types of attachments. Traffic lights and other signs without traffic functionality are not included in the Saunalahti test site. The main error is from mis-recognising traffic signs as other signs. This is because there are significant differences in the radiometric attributes between the Enschede test site and Saunalahti test site.

4.3. Analysis

We analyse the experimental results with three points. The reason why the RF classifier outperforms the SVM classifier is first investigated. We explain why the interpretation of some types of attachments only achieves low accuracy. To confirm the effectiveness of our designed features, we calculate their importance and analyse the reason.

The experimental results indicate that the random forest classifier performs the best and the GMM classifier performs the worst. SVM performs worse than random forest. Random forest is a combination of weak learners, and SVM recognises two classes by maximising the margin. In the SVM classifier, the backward elimination is performed to select the most distinctive features. Backward elimination starts with taking all features and removing each feature in turn. With each feature removed in turn, the CV accuracy is evaluated. The model with the lowest CV accuracy is discarded, and the corresponding feature is eliminated from the set of all features. This process is repeated until only one feature is left. As a result, an elimination curve is obtained in which the CV accuracy after each elimination is plotted as a function of the elimination round. The feature set with the highest CV accuracy is chosen as optimal. The starting point of backward elimination is actually in the right side of Table 7, where all features are used. In the three

Table 3
The confusion matrix of results in Enschede dataset.

	Street light	Traffic sign	Street sign	Traffic light	Connected traffic signs	Other signs	Other attachments	Correctness
Street light	148	0	0	0	0	1	1	98.7%
Traffic sign	0	43	10	0	4	9	1	64.2%
Street sign	0	9	41	0	0	1	0	80.4%
Traffic light	0	1	0	17	0	6	1	68.0%
Connected traffic signs	0	3	3	0	13	0	0	68.4%
Other signs	0	7	0	1	2	13	3	50.0%
Other attachments	1	3	2	5	1	9	84	80.0%
Completeness	99.3%	65.2%	73.2%	73.9%	65.0%	33.3%	93.3%	

Table 4
The confusion matrix of results in Saunalahti winter dataset.

	Street light	Traffic sign	Street sign	Connected traffic signs	Other attachments	Correctness
Street light	90	0	0	0	5	94.7%
Traffic sign	0	83	1	4	3	91.2%
Street sign	0	2	4	1	1	50.0%
Connected traffic signs	0	3	0	15	0	83.3%
Other attachments	2	4	0	1	144	95.4%
Completeness	97.8%	90.2%	80%	71.4%	94.1%	

Table 5
The confusion matrix of results in Saunalahti spring dataset.

	Street light	Traffic sign	Street sign	Connected traffic signs	Other attachments	Correctness
Street light	119	0	0	1	8	93.0%
Traffic sign	0	106	4	6	3	89.1%
Street sign	0	1	13	0	1	86.7%
Connected traffic signs	0	0	0	20	0	100.0%
Other attachments	2	1	0	2	205	97.6%
Completeness	98.3%	98.1%	76.5%	69.0%	94.5%	

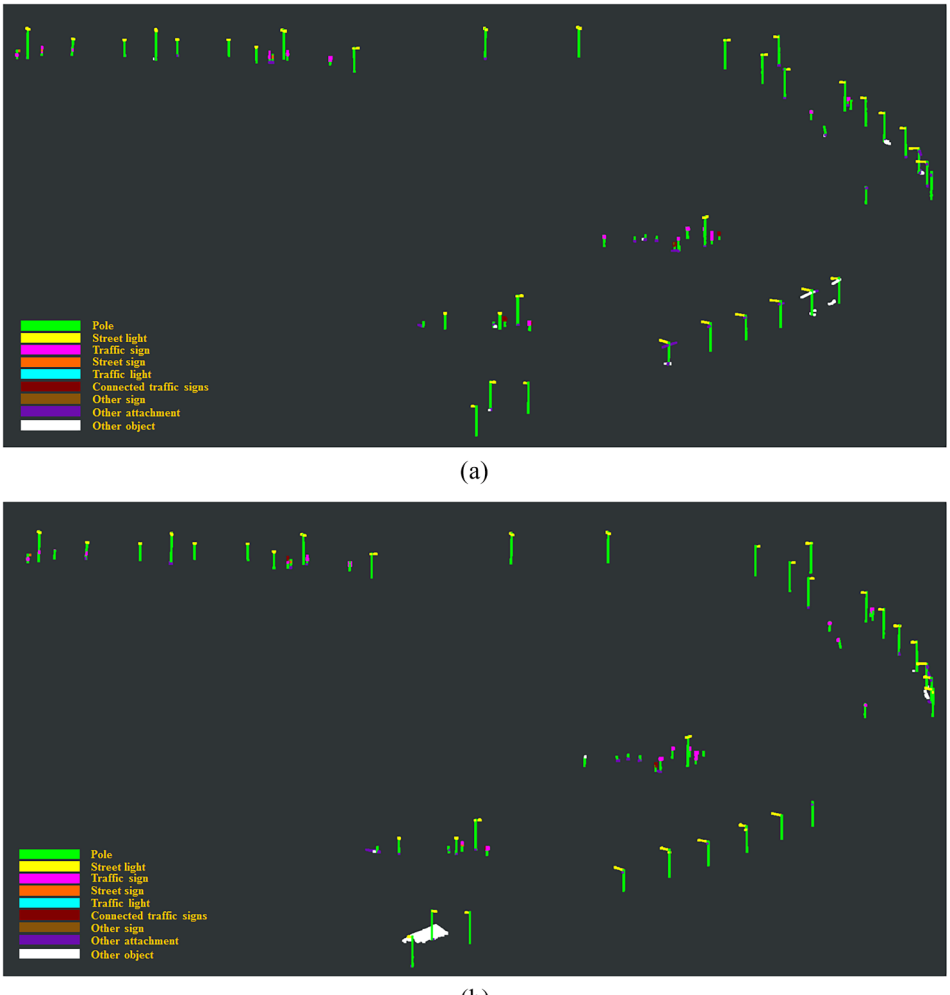
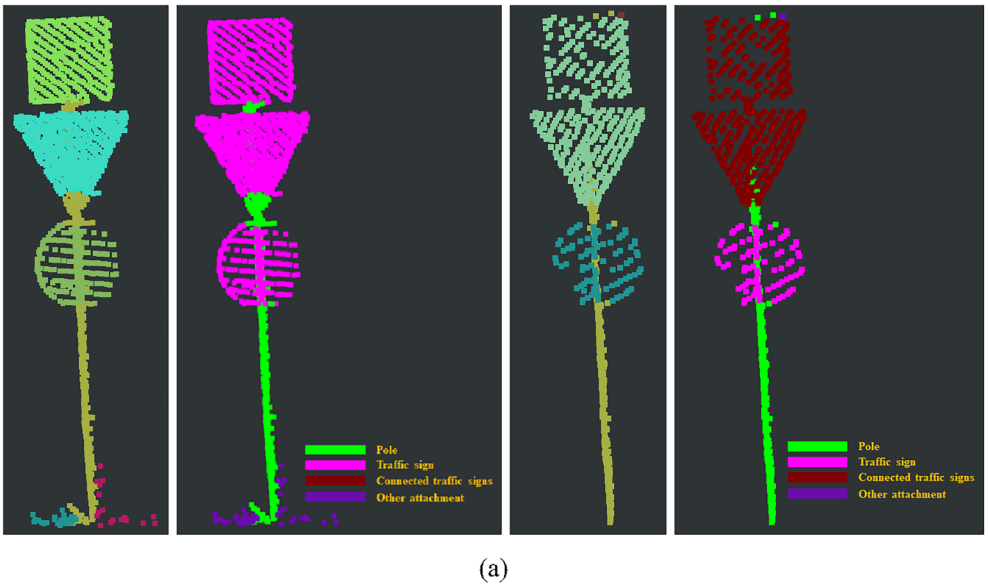
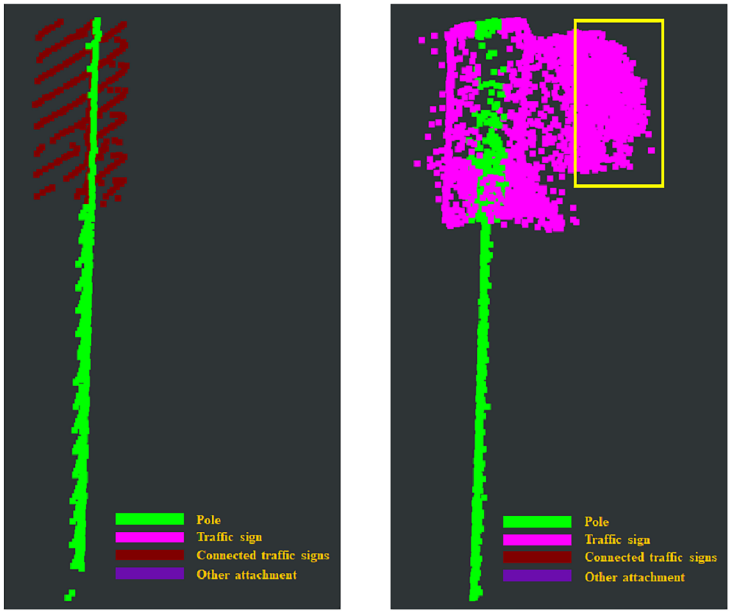


Fig. 14. The interpretation of same road furniture in two epochs in Saunalahti. The interpretation of corresponding road furniture in (a) Saunalahti winter dataset and (b) Saunalahti spring dataset.

test datasets, within the first eleven most important features all the five types of features (size features S, eigenvalue-based features E, radiometry features R, angle features A, relative height features H) are present. Table 7 illustrates the maximum performance of SVM can be reached with only eleven features. The performance of SVM classifier is not improved by adding extra features. This indicates that the performance of SVM classifier converge after using the most distinctive features (the first eleven features), that is, it is difficult for SVM to use



(a)



(b)

Fig. 15. (a) The decomposition and interpretation of road furniture and (b) point clouds with stray points. The left two figures in (a) and the left one figure in (b) are from Saunalahti winter dataset. The right two figures in (a) and the right figure in (b) are from Saunalahti spring dataset.

Table 6
The performance of transferability of trained models.

Model transferring	Random forest
Saunalahti winter training + Saunalahti spring testing	92.5%
Saunalahti spring training + Saunalahti winter testing	91.2%
Enschede training + Saunalahti winter testing	65.8%
Enschede training + Saunalahti spring testing	69.7%

designed features effectively. The reason that the accuracy does not improve after the first eleven can be that the rest features (or at least part of them) correlate strongly with (some of) the first eleven ones. Then they are redundant and may not improve the accuracy (Guyon and Elisseeff, 2003).

Compared to the random forest model, the knowledge-driven method performs much worse. The knowledge-driven method was based on generic rules in combination with a few features, whereas random forest took advantage of the distribution of features to construct many weak learners. The reason is that it is difficult to use a few generic rules with a rather limited combination of features to differentiate similar objects. As shown in Fig. 16, traffic signs are not correctly classified by the knowledge-driven method, whereas random forest classifier predicts the label of traffic sign correctly by combining the weak constraints of designed features. The combination of weak learners is similar to a combination of many manually designed rules. Via these weak learners, the excellent performance of the classification of attachments is achieved. However, it is difficult for a knowledge-driven method to use as many rules or learners as random forest to perform the classification. The knowledge-driven method does not need that much training data. Therefore, it is suitable to solve classification problems with few classes and limited data, or to make the initial training data, which is manually corrected and improved.

Table 7
Feature selection of SVM in three datasets.

Test datasets	The first eleven features					Convergence accuracy	The rest of features	Accuracy with all features
	H	S	E	A	R		H, S, E, A, R	
Enschede	H_{CT}, H_{BT}, H_{TT},	S_H	E_E, E_{LP}, E_L, e_3,	A_{VA}	R_{MA}, R_{MD}	75.6%	E_S, E_P, E_O, S_{D-H}, E_A, S_W, H_{TB}, H_{BB}, e_1, S_V, S_D, S_L, H_{CB}, e_2, A_{PA}	75.6%
Saunalahti winter	H_{TB}, H_{CB}, H_{BT}	S_H	e_1, E_O, e_2, e_3, E_{LP}	A_{VA}	R_{MA}	82.0%	H_{BB}, E_S, E_E, H_{TT}, R_{MD}, E_P, S_W, H_{CT}, E_A, A_{PA}, S_V, S_{D-H}, S_D, S_L, E_L	83.3%
Saunalahti spring	H_{BT}, H_{BB}, H_{TB},	S_{D-H}, S_W	E_E, E_O, E_{LP}, E_L	A_{VA}	R_{MD}	83.2%	R_{MA}, E_S, E_P, S_D, e_2, H_{CT}, H_{TT}, H_{CB}, S_H, e_3, E_A, e_1, S_V, S_L, A_{PA}	83.8%

Fig. 16. The result of the interpretation of road furniture from the knowledge-driven method (left) and the random forest classifier (right).

In the classification of attachments, the accuracy of the recognition of street light is the highest, higher than 90%. In contrast, the recognition rate of street signs, connected traffic signs and other signs is low, less than 80% in Saunalahti test site. One reason is that there are not many training samples and testing samples. For instance, in Saunalahti spring dataset, there are only 5 street signs in the training dataset and 17 street signs in testing dataset. By contrast, there are 58 street lights in the training dataset and 121 street lights in the testing dataset. The classifier can be biased to the class with more training samples. Another reason is that the designed features of other signs and traffic signs are quite similar in structural view. For example, both bus schedule plates and traffic signs are in drivers' view and perpendicular to the principal direction of poles. When manually making ground truth dataset, colour imagery was used to aid the labelling. It is difficult to distinguish the classes well from point clouds only.

In order to verify the effectiveness of the designed features, we also explore the feature importance in random forest. The importance of features in three test sites is as shown in Fig. 17a–c. In Fig. 17a, these bins with respective colour represent relative height features (blue), size features (red), eigenvalue-based features (green), relative angle features (cyan) and radiometric features (yellow). Radiometric features play the most significant role in all three datasets because they discriminate traffic functional signs and attachments. Relative height and size features also play an important role in the interpretation of road furniture in all three datasets. This is because there are large differences in relative height and size between different classes. Relative height features play the most significant role in the Saunalahti spring dataset. Point clouds of signs are rather noisy as aforementioned, and the calculation of other features is not reliable, especially the calculation of the normal direction of signs. That is the reason why relative height features are so important and the importance of relative angles features is low. For instance, street lights are often mounted above a certain height, and signs should be within drivers' view field. The calculation of relative height features is less affected by stray points in the point clouds of attachments. For instance, the most important relative height feature is relative height between the bottom position of an attachment and the bottom position of its connected pole. The calculation of this feature is relevant to the extraction of poles, which is robust and less affected by the stray points around attachments. In contrast, the calculation of other features (size features, eigenvalue-based features and relative angle features) is more sensitive to the noise level of point clouds of attachments. When there are noisy points surrounding them, the calculation of these features is unreliable. For instance, the size features are not authentic because of the surrounding noisy points of attachments. Therefore, the calculation of designed relative height features is robust even in noisy point clouds. Compared to other features, the angle features are not important because all signs are mounted under the same angle. Hence, the angle features cannot discriminate different types of signs. In the Saunalahti winter dataset, there is a lower level of noisy points. The calculation of size and geometric features is reliable, which contributes more to the random forest classifier. It leads to lower importance of other features such as the relative height features. Compared to the importance of relative height features in other datasets, relative height features in the Saunalahti winter dataset are less important (Fig. 17b).

5. Conclusion

We draw the conclusions as four parts: the contribution of this paper, the comparison between different machine learning classifiers,

the impact of feature importance and the reliability and transferability of our framework.

In this paper, we first decompose a point cloud of detected road furniture to separate poles and their attachments. Secondly, we generate features on attachments, pole and between attachment and poles in order to interpret road furniture. With the decomposition processing, our framework is able to interpret pole-like road furniture at a more detailed level. If there is no decomposition, complex pole-like road furniture with multiple classes cannot be discriminated reliably. Thus, features of attachments cannot be reliably extracted, and these

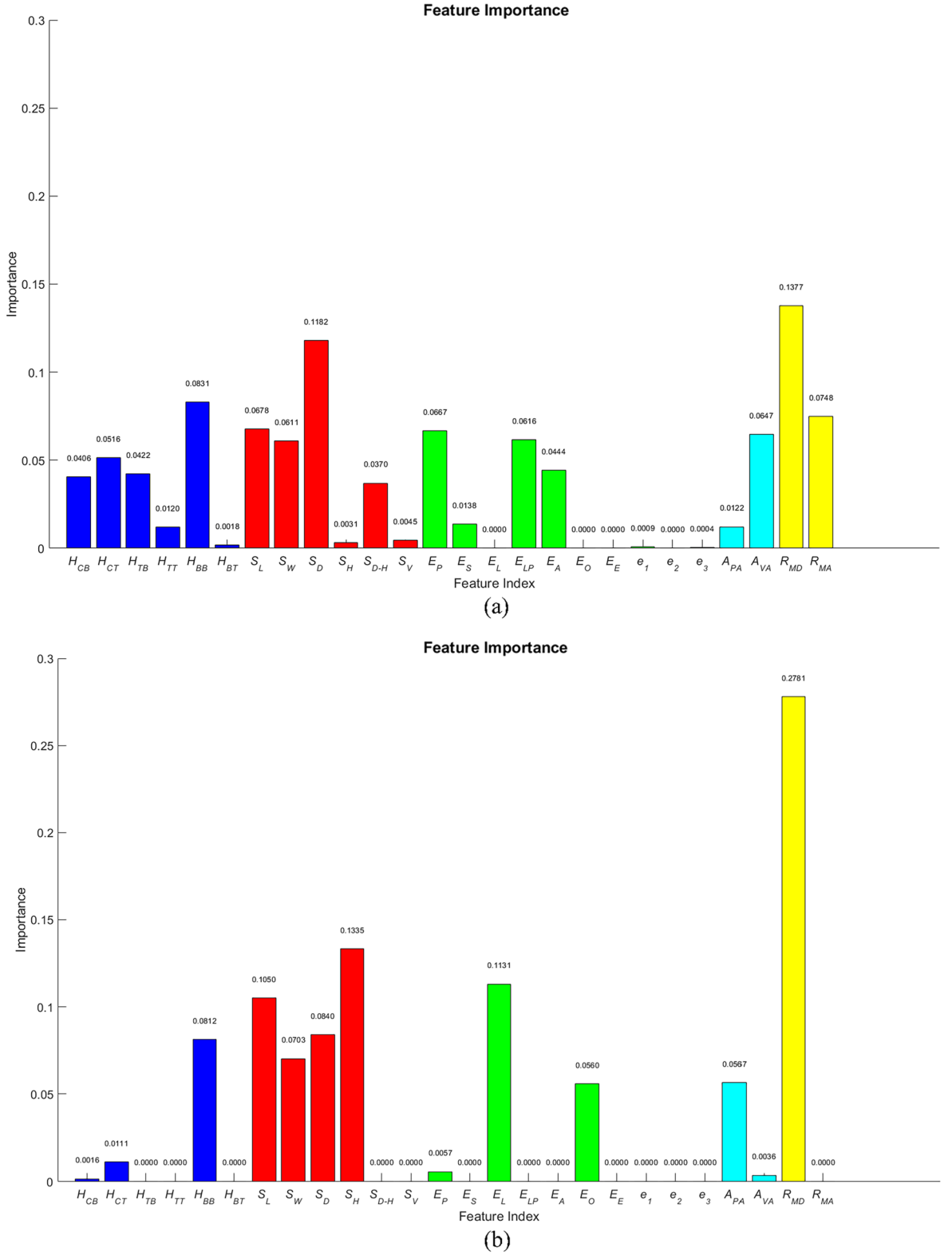
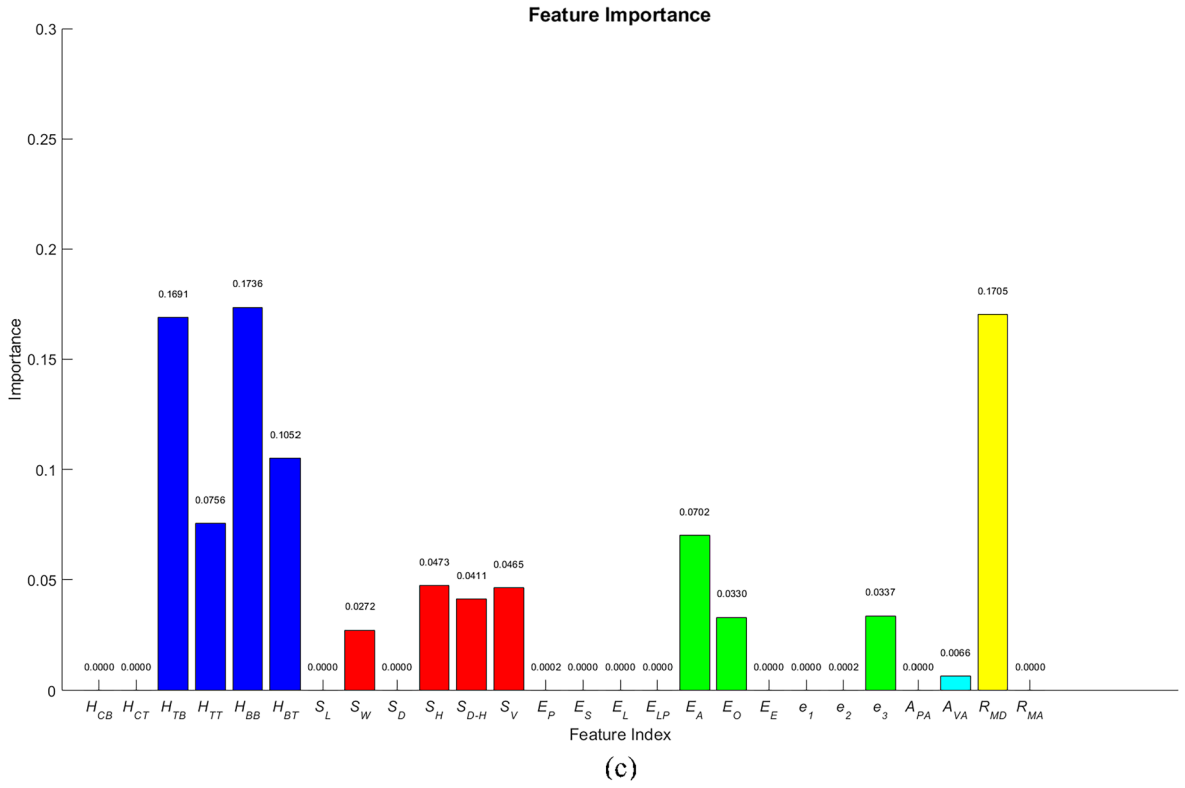


Fig. 17. Feature importance in random forest in (a) Enschede dataset, (b) Saunalahti winter dataset and (c) Saunalahti spring dataset, blue bars represent relative height features, size features are denoted by red bins, green bins indicate eigenvalue-based features, cyan bins are relative angle features, radiometric features are represented by yellow bars. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



(c)

Fig. 17. (*continued*)

attachments cannot be classified correctly. The decomposition significantly affects the interpretation stage. More than 68.0% attachments in two test sites cannot be classified correctly by previous frameworks (Yang et al., 2015; Yu et al., 2016; Wang et al., 2017). Our new framework represents a detailed interpretation of pole-like road furniture and presents the state of the art of pole-like road furniture interpretation at a functionality level, which is of great significance for precise mapping and autonomous driving. The research shows great promises for 3D precise mapping in urban road environments.

Of the four machine learning classifiers (random forest, SVM, GMM and naïve Bayes), random forest outperforms the rest of the classifiers. When there are attachments with features with small differences but with potential rules, it is difficult to distinguish them by using SVM. Even though a non-linear kernel function is able to enhance the performance of SVM classifier, a limited number of hyperplanes might still not be able to cover a large number of weak rules of random forest leading to discrimination of different classes. However, with weak learners trained in random forest with the bagging strategy such attachments can be differentiated. When the difference between features is not distinctive enough but still with potential weak rules, random forest might be a good choice.

The analysis of feature importance indicates that the designed contextual features and radiometric features play an important role in the interpretation of road furniture. The relative height features play the most important role in the point clouds which contains numerous stray points. By using relative height features, our framework is not strongly affected by stray points.

The difference of performance metrics of the interpretation of corresponding road furniture between two temporal datasets is not large, which confirms the reliability of our proposed framework. When applying the trained model of one epoch dataset to another epoch dataset, the achieved accuracy is higher than 90%. The transferability of our proposed framework between temporal datasets collected in the same area therefore is high, which proves the great potentiality of our framework applied to change detection and map updating.

Acknowledgements

We acknowledge the financial support from China Scholarship Council, the Academy of Finland thematic project “Integration of large multisource point cloud and image datasets for adaptive map updating” (No. 295047) and Centre of Excellence in Laser Scanning Research projects (No. 307362, 292735).

References

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York, NY, USA.

Boulch, A., Le Saux, B., Audebert, N., 2017, April. Unstructured Point Cloud Semantic Labeling Using Deep Segmentation Networks. In: 3DOR.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Bremer, M., Wichmann, V., Rutzinger, M., 2013. Eigenvalue and graph-based object extraction from mobile laser scanning point clouds. ISPRS Annals Photogram., Remote Sens. SpatialInform. Sci. Vol. II-5/W2, 55–60.

Brenner, C., 2009. Extraction of Features From Mobile Laser Scanning Data for Future Driver Assistance Systems. Springer, Hannover, pp. 25–42.

Cabo, C., Ordoñez, C., García-Cortés, S., Martínez, J., 2014. An algorithm for automatic detection of pole-like street furniture objects from Mobile Laser Scanner point clouds. ISPRS J. Photogram. Remote Sens. 87, 47–56.

Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST) 2 (3), 27.

Criminisi, A., Shotton, J., Konukoglu, E., 2012. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Found. Trends Comput. Graph. Vision 7 (2–3), 81–227.

El-Halawany, S.I., Lichti, D.D., 2013. Detecting road poles from mobile terrestrial laser scanning data. GIScience Remote Sens. 50 (6), 704–722.

Fukano, K., Masuda, H., 2015. Detection and classification of pole-like objects from mobile mapping data. ISPRS Annals of Photogram., Remote Sens. Spatial Inform. Sci. 1, 57–64.

Golovinskiy, A., Kim, V.G., Funkhouser, T., 2009. Shape-based recognition of 3D point clouds in urban environments. In: IEEE International Conference on Computer Vision, Kyoto, Japan, pp. 2154–2161.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3 (Mar), 1157–1182.

Hackel, T., Wegner, J.D., Schindler, K., 2016. Fast semantic segmentation of 3D point clouds with strongly varying density. ISPRS Annals Photogram., Remote Sens. Spatial Inform. Sci. Vol. III-3, 177–184.

Hsu, C.W., Chang, C.C., Lin, C.J., 2003. A Practical Guide to Support Vector Classification.

[Online]. Available: < http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf > .
Hsu, C.W., Lin, C.J., 2002. A comparison of methods for multiclass support vector machines. IEEE Trans. Neural Networks 13 (2), 415–425.
Huang, J., You, S., 2015. Pole-like object detection and classification from urban point clouds. In: IEEE International Conference on Robotics and Automation, Seattle, Washington USA, pp. 3032–3038.
Huang, J., You, S., 2016. December. Point cloud labeling using 3d convolutional neural network. In: Pattern Recognition (ICPR), 2016 23rd International Conference on. IEEE, pp. 2670–2675.
Landrieu, L., Simonovsky, M., 2017. Large-scale point cloud semantic segmentation with superpoint graphs. arXiv preprint arXiv:1711.09869.
Lehtomäki, M., Jaakkola, A., Hyyppä, J., Kukko, A., Kaartinen, H., 2010. Detection of vertical pole-like objects in a road environment using vehicle-based laser scanning data. Remote Sens. 2 (3), 641–664.
Lehtomäki, M., Jaakkola, A., Hyyppä, J., Lampinen, J., Kaartinen, H., Kukko, A., Puttonen, E., Hyyppä, H., 2016. Object classification and recognition from mobile laser scanning point clouds in a road environment. IEEE Trans. Geosci. Remote Sens. 54 (2), 1226–1239.
Li, D., Oude Elberink, S., 2013. Optimizing detection of road furniture (pole-like objects) in mobile laser scanner data. ISPRS Annals Photogram., Remote Sens. Spatial Inform. Sci. Vol. II-5/W2, 163–168.
Li, F., Oude Elberink, S., Vosselman, G., 2017. Semantic labelling of road furniture in mobile laser scanning data. Int. Arch. Photogram., Remote Sens. Spatial Inform. Sci., vol. 42.
Li, F., Oude Elberink, S., Vosselman, G., 2018a. Pole-like road furniture detection and decomposition in mobile laser scanning data based on spatial relations. Remote Sens. 10 (4), 531.
Li, F., Lehtomäki, M., Oude Elberink, S., Vosselman, G., Puttonen, E., Kukko, A., Hyyppä, J., 2018b. Pole-like road furniture detection in sparse and unevenly distributed mobile laser scanning data. ISPRS Annals Photogram., Remote Sens. Spatial Inform. Sci. 4 (2).
Li, Y., Bu, R., Sun, M. and Chen, B., 2018c. PointCNN. arXiv preprint arXiv:1801.07791.
Louppe, G., Wehenkel, L., Sutera, A., Geurts, P., 2013. Understanding variable importances in forests of randomized trees. In: Advances in Neural Information Processing Systems, pp. 431–439.
Ma, L., Li, Y., Li, J., Wang, C., Wang, R., Chapman, M., 2018. Mobile laser scanned point-clouds for road object detection and extraction: a review. Remote Sens. 10 (10), 1531.
Maturana, D., Scherer, S., 2015, September. Voxnet: A 3d convolutional neural network for real-time object recognition. In: Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on IEEE, pp. 922–928.
Munoz, D., Bagnell, J.A., Vandapel, N., Hebert, M., 2009. Contextual classification with functional max-margin Markov networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 975–982.
Murphy, K.P., 2012. Machine learning: a probabilistic perspective.
Oude Elberink, S., Kemboi, B., 2014. User-assisted object detection by segment based similarity measures in mobile laser scanner data. Int. Arch. Photogram., Remote Sens. Spatial Inform. Sci. Vol. XL-3, 239–246.
Pu, S., Rutzinger, M., Vosselman, G., Oude Elberink, S., 2011. Recognizing basic structures from mobile laser scanning data for road inventory studies. ISPRS J. Photogram. Remote Sens. 66 (6), S28–S39.
Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. Pointnet: deep learning on point sets for 3d classification and segmentation. Proc. Comput. Vision Pattern Recogn. (CVPR), IEEE 1 (2), 4.
Riegler, G., Ulusoy, A.O., Geiger, A., 2017, July. Octnet: Learning deep 3d representations at high resolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 3.
Riveiro, B., Díaz-Vilariño, L., Conde-Carnero, B., Soilán, M., Arias, P., 2016. Automatic segmentation and shape-based classification of retro-reflective traffic signs from mobile lidar data. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 9 (1), 295–303.
Kohavi, R., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In: Proc. 14th International Joint Conference on Artificial Intelligence, Montreal, Quebec, Canada, pp. 1137–1143.
Ross, S., Munoz, D., Hebert, M., Bagnell, J.A., 2011. Learning message-passing inference machines for structured prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2737–2744.
Soilán, M., Riveiro, B., Martínez-Sánchez, J., Arias, P., 2016. Traffic sign detection in MLS acquired point clouds for geometric and image-based semantic inventory. ISPRS J. Photogram. Remote Sens. 114, 92–101.
Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E., 2015. Multi-view convolutional neural networks for 3d shape recognition. In: In Proceedings of the IEEE International Conference on Computer Vision, pp. 945–953.
Su, H., Jampani, V., Sun, D., Maji, S., Kalogerakis, E., Yang, M.H., Kautz, J., 2018. Splatnet: Sparse lattice networks for point cloud processing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2530–2539.
Tombari, F., Fioraio, N., Cavallari, T., Salti, S., Petrelli, A., Di Stefano, L., 2014. Automatic detection of pole-like structures in 3d urban environments. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4922–4929.
Velizhev, A., Shapovalov, R., Schindler, K., 2012. Implicit shape models for object detection in 3D point clouds. ISPRS Annals Photogram., Remote Sens. Spatial Inform. Sci. Vol. I-3, 179–184.
Vosselman, G., Gorte, B.G., Sithole, G., Rabbani, T., 2004. Recognising structure in laser scanner point clouds. Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci. Vol. XXXVI-8/W2, 33–38.
Wang, J., Lindenbergh, R., Menenti, M., 2017. SigVox–A 3D feature matching algorithm for automatic street object recognition in mobile laser scanning point clouds. ISPRS J. Photogram. Remote Sens. 128, 111–129.
Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS J. Photogram. Remote Sens. 105, 286–304.
Xiong, X., Munoz, D., Bagnell, J.A., Hebert, M., 2011. 3-D scene analysis via sequenced predictions over points and regions. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 2609–2616.
Yang, B., Dong, Z., 2013. A shape-based segmentation method for mobile laser scanning point clouds. ISPRS J. Photogram. Remote Sens. 81, 19–30.
Yang, B., Dong, Z., Zhao, G., Dai, W., 2015. Hierarchical extraction of urban objects from mobile laser scanning data. ISPRS J. Photogram. Remote Sens. 99, 45–57.
Yokoyama, H., Date, H., Kanai, S., Takeda, H., 2013. Detection and classification of pole-like objects from mobile laser scanning data of urban environments. Int. J. CAD/CAM 13 (2), 31–40.
Yu, Y., Li, J., Guan, H., Wang, C., Yu, J., 2015. Semiautomated extraction of street light poles from mobile LiDAR point-clouds. IEEE Trans. Geosci. Remote Sens. 53 (3), 1374–1386.
Yu, Y., Li, J., Wen, C., Guan, H., Luo, H., Wang, C., 2016. Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data. ISPRS J. Photogram. Remote Sens. 113, 106–123.