Robust Traffic-Sign Detection and Classification Using Mobile LiDAR Data With Digital Images

Haiyan Guan, *Member, IEEE*, Wanqian Yan, Yongtao Yu, *Member, IEEE*, Liang Zhong, and Dilong Li

***Abstract*—This study aims at building a robust method for detecting and classifying traffic signs from mobile LiDAR point clouds and digital images. First, this method detects traffic signs from mobile LiDAR point clouds with regard to a prior knowledge of road width, pole height, reflectance, geometrical structure, and traffic-sign size. Then, traffic-sign images are segmented by projecting the detected traffic-sign points onto the digital images. Afterward, the segmented traffic-sign images are normalized for automatic classification with a given image size. Finally, a traffic-sign classifier is proposed based on a supervised Gaussian–Bernoulli deep Boltzmann machine model. We evaluated the proposed method using datasets acquired by a RIEGL VMX-450 system. The traffic-sign detection accuracy of 86.8% was achieved; through parameter sensitivity analysis, the overall performance of traffic-sign classification achieved a recognition rate of 93.3%. The computational performance showed that our method provides a promising solution to traffic-sign detection and classification using mobile LiDAR point clouds and digital images.**

***Index Terms*—Digital images, deep learning, geometrical features, intensity, mobile LiDAR point clouds, traffic signs.**

I. Introduction

TRAFFIC signs play an important role in road transportation systems because they provide useful and vital road information and instruction to road users [1]. For example, to regulate the traffic safety, variable speed limitations, informational signs, and directional signs are placed along the road according to the environmental conditions and traffic situations of the road. Therefore, rapidly updating traffic signs is essential for transportation agencies to manage and monitor the status and usability of traffic signs [2], [3].

Current traffic sign detection and recognition systems are based mainly on digital images and videos [4]. To warn and guide drivers, traffic signs, well defined by highly contrasting colors (e.g., red, blue, yellow, and white), can be distinguished easily from a complex environment. Additionally, transportation systems mainly use circular, triangular, rectangular, square, octagonal, and pentagonal shapes of traffic signs to regulate the traffic. The combination of shapes and colors represents traffic-sign categories, such as prohibition, danger, obligation, and warning. Therefore, based on image features derived from color and shape information, traffic signs are usually described by various feature descriptors, including the following: scale-invariant feature transform, wavelet decomposition, histograms of oriented gradients, local binary pattern, and correlating Fourier descriptors [1], [5]–[7]. Image- or video-based traffic-sign detection and recognition systems generally have been developed by applying the features to a variety of classifiers and object-classification algorithms, such as the following: support vector machine, random forest, neural network, convolutional neural network (CNN), decision fusion and reasoning module, and template-matching-based methods [8]–[14]. However, these image- or video-based systems suffer from the following limitations:

1) weather conditions (e.g., fog and rain), affecting the visibility of traffic signs,
2) shadows, caused by other adjacent objects or different illumination levels,
3) light condition, and
4) sign condition, including color fading, deformation, and similarity of signs [1].

Mobile laser scanning or mobile LiDAR systems, integrated with laser scanners, cameras, and a set of navigation sensors, rapidly acquire highly dense and accurate three-dimensional (3-D) point clouds; therefore, mobile LiDAR technology is being used at an increasing rate for transportation-related surveys [15]–[18]. Most existing traffic-sign detection methods for mobile LiDAR point clouds are based basically on prior knowledge, such as shape, position, and reflectance [19]. Yu *et al.* [20] detected traffic signs based on 3-D points via a bag-of-visual-phrases representation. The traffic sign detection task has proven to be effective and reliable. However, this method has two aspects to be improved: 1) the use of bag-of-visual-phrases representation increases the computational complexity at traffic

Manuscript received September 2, 2017; revised December 8, 2017 and February 12, 2018; accepted February 22, 2018. Date of publication March 25, 2018; date of current version May 1, 2018. This work was supported in part by the Natural Science Foundation of Jiangsu Province under Grant BK20151524 and Grant BK20160427, in part by the National Natural Science Foundation of China under Grant 41501501, Grant 61603146, and Grant 41671454, in part by the Natural Science Research in Colleges and Universities of Jiangsu Province under Grant 16KJB520006, in part by the Science and Technology Project of Huaian City under Grant HAG201602, and in part by the Jiangsu "Shuangchuang" project. *(Corresponding author: Haiyan Guan.)*

H. Guan and W. Yan are with the College of Geography and Remote Sensing, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: guanhy.nj@nuist.edu.cn; ywq_nuist@163.com).

Y. Yu is with the Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 223003, China (e-mail: allennessy.yu@gmail.com).

L. Zhong is with Changjiang Spatial Information Technology Engineering Co., Ltd., Wuhan 410010, China (e-mail: bluewind20000@163.com).

D. Li is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430072, China, and also with the China Transport Telecommunications and Information Center, Beijing 100011, China (e-mail: scholar.dll@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2018.2810143

sign detection stage; 2) the performance of individual traffic sign detection relies on effective segmentation results, which might affect the detection performance in complex road environment. To detect traffic signs placed closely to road boundaries, Chen *et al.* [21] proposed a processing chain of cross-section analysis, individual object segmentation, and linear structure inference. Given that traffic signs are presented as linearly shaped structures, Yokoyama *et al*. [22] detected traffic signs by 1) applying Laplacian smoothing to point cloud segments for suppressing measurement noise and point distribution bias, and 2) performing principal component analysis on the smoothed segments. Yu *et al*. [23] presented a robust 3-D object matching framework, which is able to detect traffic signs hidden in trees or of varying shapes and completeness. In addition, Hough forest methods were also developed for traffic sign detection [24]. Pu *et al*. [25] detected traffic signs based on percentile analysis and planar shape analysis. To guide road users, traffic signs are made from highly reflective materials. Thus, compared to other linearly shaped objects, such as light poles, traffic signs exhibit high retro-reflectivity (in the form of intensity) in mobile LiDAR point clouds [3], [26]. Based on this characteristic, traffic signs are distinguished from other objects by established algorithms, such as random sampling consensus and template-driven methods [27], [28].

Currently, a mobile mapping system is an integration of multiple sensors, including laser scanners and digital cameras [29]. Point clouds provide accurate geometric and localization information of objects; whereas, digital images provide detailed texture and content information of the objects. Therefore, by fusing images and point clouds, a mobile LiDAR system provides a promising solution to traffic-sign detection (based on LiDAR point clouds) and classification (based on digital images).

Recently, deep learning techniques have become attractive for their superior performance in learning hierarchical features from high-dimensional unlabeled data. By learning multilevel feature representations, deep learning models have proved effective in many fields such as speech recognition, natural language processing, and imaging processing [30]. Deep Boltzmann machines (DBMs) [31] are an important breakthrough in the requirement for powerful deep feature representation models. Han *et al*. [30], to effectively improve object detection performance in optical remote sensing images, built a high-level feature representation DBM to capture the spatial and structural information encoded in the low-level and middle-level features. Leng *et al.* [32] described a high-level abstraction representation for 3-D model recognition. Although the DBM and its variants have shown great potentials in presenting various feature representation models in a range of tasks, most of these models are learned in an unsupervised way (without using class labels) [33], decreasing their discriminative feature learning abilities in some recognition tasks. A Gaussian–Bernoulli DBM (GB-DBM) [31], which contains a continuous representation, instead of binary representation, in the visible layer of the DBM as the input, is effective for image processing tasks. Similarly, to guide the GB-DBM model to encode class-oriented feature representations, a modified version of the GB-DBM trained with extra class restricted information (i.e., supervised GB-DBM) is more powerful for classification applications.

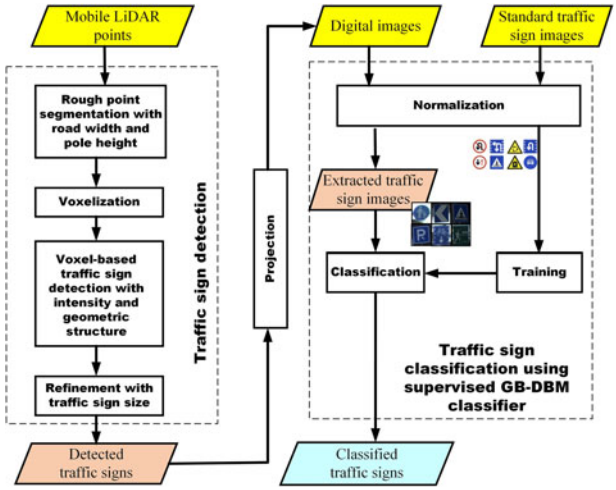


Fig. 1. Workflow of the proposed traffic-sign detection and classification method.

In this paper, using both mobile LiDAR point clouds and digital images, we propose a traffic-sign detection and classification system. The remainder of the paper is organized as follows. Section II presents an overview of the proposed traffic-sign detection and classification system. Section III details traffic-sign detection in mobile LiDAR point clouds. Section IV presents traffic-sign classification on digital images. Section V reports and discusses the results obtained using a set of mobile LiDAR point clouds and digital images. Section VI presents concluding remarks.

II. System Overview

Our proposed system detects and classifies traffic signs in two stages: 1) traffic sign interest regions are determined from mobile LiDAR data; 2) traffic signs are classified from digital images collected by imaging sensors in the same mobile LiDAR system. The traffic-sign training samples in our experiments corresponded to those used by the Ministry of Transport, China. As depicted in Fig. 1, our procedure has the following stages:

1) Traffic-sign detection in mobile LiDAR data: First, a prior knowledge of pole height and road width is used to remove the majority of points from mobile LiDAR data. Then, the remaining points are voxelized to detect traffic sign interest regions based on intensity information and geometrical structures. Finally, by using traffic-sign size information, a refinement procedure is performed to obtain the traffic-sign detection results.
2) Traffic-sign classification on digital images: First, traffic-sign images are determined by projecting the detected traffic-sign points onto their corresponding digital images. Then, via a supervised GB-DBM using standard traffic-sign samples downloaded from the website of the Ministry of Transport, China, the traffic-sign images are classified into their specific categories.

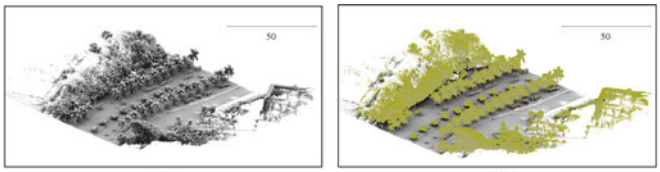


Fig. 2. Mobile LiDAR data sample. (a) Raw data (rendered by grey). (b) The filtered results with the given pole height (Γ_{PH}) and road width (Γ_{RW}) (rendered by olive).

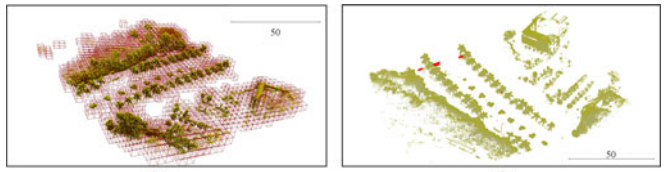


Fig. 3. Mobile LiDAR data sample. (a) Voxelization. (b) The extracted results (traffic signs: red).

The two stages are described in detail in the next two sections.

III. Traffic Sign Detection in Mobile LiDAR Data

The goal of this stage is to extract traffic-sign interest regions from mobile LiDAR data. Traffic signs usually stand out from their environment due to their characteristics, such as shape, intensity, and color. Mobile LiDAR data provide accurate positional and intensity information of traffic signs; therefore, to facilitate traffic-sign detection, with the features and a prior knowledge of traffic signs, an analysis is performed by

1) Pole height (A_{PH}): Along the roadsides, some traffic signs are directly installed at the top of poles, while some traffic signs are hung over the roads by beams supported by poles. Thus, although differing from country to country, traffic signs should be installed or placed at a certain height for drivers or pedestrians to easily capture road traffic information.
2) Road width (A_{RW}): Traffic signs are normally located along or hang over the roadsides. The width of surveyed roads provides helpful prior knowledge for roughly delineating traffic-sign regions. Besides points and digital images, a mobile LiDAR system normally provides vehicle trajectory data along the direction in which the vehicle is moving. According to the vehicle trajectory data, road width can be roughly estimated.
3) Intensity (A_I): To regulate transportation flow and safely guide drivers along the roads, traffic-sign plates are usually painted with special high-reflectance materials, so that the intensity of the traffic signs is higher than that of their surroundings.
4) Geometrical structure (A_G): Although traffic signs have different shapes (rectangle, circle, hexagon, and triangle) to demonstrate different regulations, they are planar structures with a certain area.
5) Traffic-sign size (A_A): To provide necessary traffic instructions for road users, traffic-sign plates have a certain area; for example, a circular "stop" sign with a radius of 600 mm and a square "parking" sign with a side of 1000 mm. The traffic-sign size is defined by the outline of a traffic sign.

Accordingly, the two features—pole height and road width—are first used to filter out the majority of ground and nonground points. Fig. 2(a) shows a small sample of mobile LiDAR data covering a road length of 20 m. Fig. 2(b) shows the filtered results when the thresholds of pole height (Γ_{PH}) and road width (Γ_{RW}) are set to 1.5 and 10 m, respectively.

Mobile LiDAR systems collect, in a short time, a large number of points along the road. Although we filter out the majority of mobile LiDAR data using the thresholds of pole height and road width, the number of remaining points is still considerable. Therefore, point-wise point processing methods are usually time-consuming and require a heavy computational burden. To reduce computational cost when processing voluminous mobile LiDAR points, we propose a voxel-based traffic-sign detection method. The entire scene is partitioned into a set of voxels of size V_s by means of an octree structure [see Fig. 3(a)]. The voxel size threshold Γ_S is empirically defined by the point density of the collected mobile LiDAR data. We set the centroid of each voxel v with the attribute set $C = \{A_I, A_G\}$. A_I is defined by the mean normalized intensity value based on the normalized intensity values of all points in the voxel. Based on the assumption that most points in a local neighborhood show a similar geometrical characteristic, the geometrical structure at the centroid of the voxel is calculated as follows.

Denote p_i $(i = 1, 2, \ldots N)$ as a LiDAR point in a voxel, p_c as the centroid of the voxel; then, the covariance matrix M is given by

$$\mathbf{M} = \frac{1}{N}\sum_{i=1}^{N}(p_i - p_c)(p_i - p_c)^T. \tag{1}$$

The eigenvalues $\lambda_i (i = 1, 2, 3)$ $(\lambda_1 \geq \lambda_2 \geq \lambda_3)$ and their corresponding eigenvectors of the covariance matrix M are calculated. Let a_l, a_p, and a_v denote the linear, planar, and volumetric geometrical features, respectively [34]. According to the calculated eigenvalues λ_i, a_l, a_p, and a_v are defined by

$$a_l = \frac{\sqrt{\lambda_1} - \sqrt{\lambda_2}}{\sqrt{\lambda_1}},\ a_p = \frac{\sqrt{\lambda_2} - \sqrt{\lambda_3}}{\sqrt{\lambda_1}},\ a_v = \frac{\sqrt{\lambda_3}}{\sqrt{\lambda_1}}. \tag{2}$$

Therefore, for a voxel, the geometrical structure A_G is determined by

$$A_G = \arg\max\{a_l, a_p, a_v\}. \tag{3}$$

Considering the attributes of intensity and geometrical structures, the voxel-based clusters classified as traffic signs are maintained if their intensity value is higher than the predefined threshold, as well as if their geometrical structures are planar [see Fig. 3(b)]. The predefined intensity threshold Γ_I can be estimated from the manually selected traffic signs in survey areas.

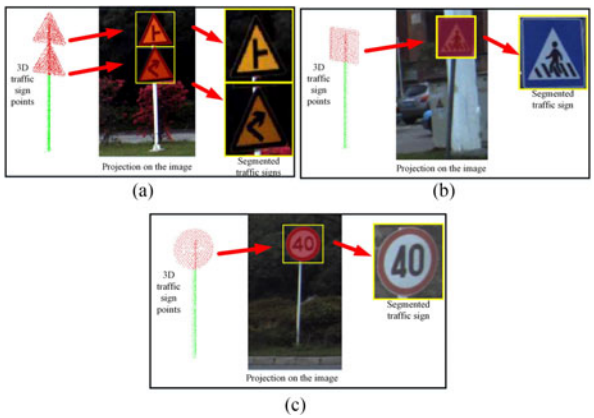


Fig. 4. Three traffic-sign samples projecting on their corresponding digital images: (a) triangular, (b) rectangular, and (c) circular traffic signs.

To further remove nontraffic-sign small clusters and noise, a refinement procedure is required. The labeled traffic sign points are first clustered into a set of clusters by a Euclidean distance clustering algorithm. The clustering distance C_{dist} is related to the parameter of voxel size (empirically, the clustering distance threshold $\Gamma_{\text{dist}} = 2 \times \Gamma_S$). The outlines for individual clusters are then obtained. Based on the outlines, their incircles are calculated. Although different traffic signs have different sizes due to their functionalities, traffic signs (e.g., for prohibition) must have a certain size for guiding road users. Roughly estimated from the representatives of traffic signs, in this study, the threshold of traffic-sign size Γ_A was constantly set to 0.2 m^2.

IV. Deep-Learning Based Traffic-Sign Classification on Digital Images

The goal of this stage is to classify traffic signs on digital images, collected by the mobile mapping system. The mobile LiDAR system captures point clouds, digital images, and vehicle trajectory data when a vehicle is traveling along a road. The collected point clouds and digital images are then geo-referenced by the vehicle trajectory data. Point clouds accurately record the positional information of traffic signs, whereas digital images provide abundant spectral information. By combining the two different data types, traffic signs are easily located and classified. After traffic signs are detected from point clouds, they are then projected onto the digital images for further traffic-sign classification. To tolerate registration errors between point clouds and digital images, we enlarge the areas of interest of the traffic signs in the digital images with a predefined toleration threshold. Fig. 4 shows three traffic-sign images selected by projecting their points onto digital images.

In this study, we propose a traffic-sign classifier based on a supervised GD-DBM model. As shown in Fig. 5(a), the training data, downloaded from the website of the Ministry of Transport, China, are trained to construct a multilayer GB-DBM model, which includes a visual, a label, and multiple hidden layers. Denote $\boldsymbol{v}_{\text{in}}$ as a vector of real-valued visible units that represents a linearly vectorized traffic-sign training sample. Denote $\boldsymbol{L}_{\text{in}}$ as a binary label vector with a "1-of-*K*" encoding pattern [31]. That is, for a vector with *K* elements, only one element is set to "1"; the others are set to "0." Denote $\boldsymbol{h}^r$ ($r = 1, 2, \ldots, N$; *N* is the number of hidden layers) as the binary hidden units in each hidden layer. Then, the energy of the joint configuration $E\{\boldsymbol{v}_{\text{in}}, \boldsymbol{L}_{\text{in}}, \boldsymbol{h}^1, \boldsymbol{h}^2, \ldots, \boldsymbol{h}^N\}$ is defined as

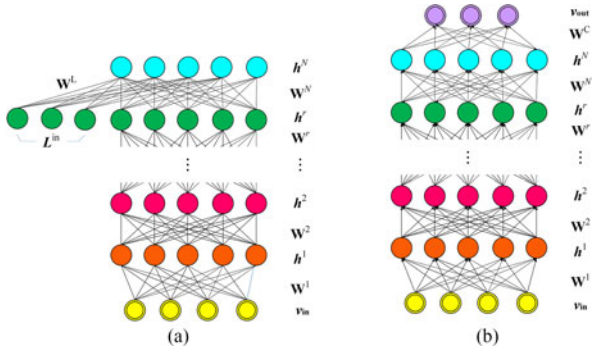


Fig. 5. Illustrations of (a) a supervised Gaussian–Bernoulli DBM model and (b) a supervised GB-DBM classifier for traffic-sign classification.

$$
\begin{aligned}
E\left(\boldsymbol{v}_{\text{in}}, \boldsymbol{L}_{\text{in}}, \boldsymbol{h}^1, \boldsymbol{h}^2, \ldots, \boldsymbol{h}^N; \boldsymbol{\theta}\right) = {} & \frac{1}{2}\sum_i \frac{v_i^2}{\sigma_i^2} - \sum_{i,j} \frac{v_i}{\sigma_i} w_{ij}^1 h_j^1 \\
& - \sum_{jm} h_j^1 w_{jm}^2 h_m^2 - \quad \cdots \\
& - \sum_{qk} h_q^{N-1} w_{qk}^N h_k^N - \sum_{nk} l_n w_{nk}^L h_k^N \\
& - \sum_i \frac{v_i}{\sigma_i} b_i - \sum_j h_j^1 a_j^1 \\
& - \sum_m h_m^2 a_m^2 - \quad \cdots \\
& - \sum_q h_q^{N-1} a_q^{N-1} - \sum_k h_k^N a_k^N
\end{aligned} \tag{4}
$$

where $\boldsymbol{\theta} = \{\mathbf{W}^1, \mathbf{W}^2, \ldots, \mathbf{W}^r, \ldots \mathbf{W}^N, \mathbf{W}^L, \boldsymbol{\sigma}^2, \boldsymbol{b}, \boldsymbol{a}^1, \boldsymbol{a}^2, \ldots, \boldsymbol{a}^N\}$ are the model parameters. $\mathbf{W}^1$ represents the visible-to-hidden symmetric interaction term; $\mathbf{W}^L$ represents the label-to-hidden symmetric interaction term; $\mathbf{W}^r$ ($r = 2, \ldots, N$) represents the *r*th hidden-to-hidden symmetric interaction terms; $\boldsymbol{\sigma}^2$ represents the variances of the visible units; $\boldsymbol{b}$ represents the biases in the visible layer; $\boldsymbol{a}^r$ ($r = 1, \ldots, N$) represents the *r*th biases in the hidden layers. The marginal distribution over the visible vector $\boldsymbol{v}_{\text{in}}$ with a label vector $\boldsymbol{L}_{\text{in}}$ takes the form in (5), shown at the bottom of this page.

$$
P\left(\boldsymbol{v}_{\text{in}}, \boldsymbol{L}_{\text{in}}; \theta\right) = \frac{\sum_{\boldsymbol{h}^1, \boldsymbol{h}^2, \cdots, \boldsymbol{h}^N} \exp\left[-E\left(\boldsymbol{v}_{\text{in}}, \boldsymbol{L}_{\text{in}}, \boldsymbol{h}^1, \boldsymbol{h}^2, \cdots, \boldsymbol{h}^N; \theta\right)\right]}{\int_{\boldsymbol{t}'_{\text{in}}} \sum_{\boldsymbol{h}^1, \boldsymbol{h}^2, \cdots, \boldsymbol{h}^N, \boldsymbol{L}'_{\text{in}}} \exp\left[-E\left(\boldsymbol{t}'_{\text{in}}, \boldsymbol{L}'_{\text{in}}, \boldsymbol{h}^1, \boldsymbol{h}^2, \cdots, \boldsymbol{h}^N; \theta\right)\right] d\boldsymbol{t}'_{\text{in}}} \tag{5}
$$

The conditional distributions over the visible, label, and multiple sets of hidden units are expressed, respectively, as

$$p\left(v_i = x|\boldsymbol{h}^1\right) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{\left(x-\sigma_i\left(\sum_j w_{ij}^1 h_j^1 + b_i\right)\right)}{2\sigma_i^2}\right) \tag{6}$$

$$p\left(l_n|\boldsymbol{h}^N\right) = \frac{\exp\left(\sum_k w_{nk}^L h_k^N\right)}{\sum_s \exp\left(\sum_k w_{sk}^L h_k^N\right)} \tag{7}$$

$$\begin{cases} p\left(h_j^1 = 1|\boldsymbol{v_{\text{in}}}, \boldsymbol{h}^2\right) = g\left(\sum_i w_{ij}^1 \frac{v_i}{\sigma_i} + \sum_m w_{jm}^2 h_m^2 + a_j^1\right) \\ p\left(h_m^2 = 1|\boldsymbol{h}^1, \boldsymbol{h}^3\right) = g\left(\sum_j w_{jm}^2 h_j^1 + \sum_k w_{mk}^3 h_k^3 + a_m^2\right) \\ \vdots \\ p\left(h_k^N = 1|\boldsymbol{h}^{N-1}, \boldsymbol{L_{\text{in}}}\right) = g\left(\sum_q w_{qk}^N h_q^{N-1} + \sum_n w_{nk}^L l_n + a_k^N\right) \end{cases} \tag{8}$$

To train the model parameters θ of the supervised GB-DBM model, we first initialize the model parameters by a greedy layer-wise pretraining algorithm. Then, we fine-tune the model parameters by an iterative joint training algorithm integrated with variational and stochastic approximation approaches [31], [35]. Next, we construct a supervised GB-DBM classifier by placing a logistic regression (LR) layer on the top of the highest hidden layer [see Fig. 5(b)] [31]. The LR layer contains a set of softmax units, each of which corresponds to a specific class. Note that, in our GB-DBM classifier, each hidden layer provides deterministic, real-valued probability estimations, rather than stochastic activities of binary features in the supervised GB-DBM model. Then, the most representative, high-level feature $\boldsymbol{h}^N$ (N is the number of hidden layers) is an input to the LR layer. Finally, the GB-DBM classifier is fine-tuned based on the standard back-propagation of error derivatives [35]. Given a visible vector $\boldsymbol{v}_{\text{in}}$, the output $\boldsymbol{v}_{\text{out}}$ of the supervised GB-DBM classifier is computed by

$$\begin{aligned} \left(\boldsymbol{h}^1\right)^T &= g\left(\frac{\boldsymbol{v_{\text{in}}^T}}{\sigma^T}\mathbf{W}^1 + \left(\boldsymbol{a}^1\right)^T\right) \\ \left(\boldsymbol{h}^2\right)^T &= g\left(\left(\boldsymbol{h}^1\right)^T\mathbf{W}^2 + \left(\boldsymbol{a}^2\right)^T\right) \\ \left(\boldsymbol{h}^3\right)^T &= g\left(\left(\boldsymbol{h}^2\right)^T\mathbf{W}^3 + \left(\boldsymbol{a}^3\right)^T\right) \\ &\vdots \\ \left(\boldsymbol{h}^N\right)^T &= g\left(\left(\boldsymbol{h}^{N-1}\right)^T\mathbf{W}^N + \left(\boldsymbol{a}^N\right)^T\right) \end{aligned} \tag{9}$$

$$v_{\text{out}}^{(n)} = \frac{\exp\left(\sum_k w_{kn}^C h_k^3\right)}{\sum_s \exp\left(\sum_k w_{ks}^C h_k^3\right)}. \tag{10}$$

TABLE I
DESCRIPTION OF THE THREE MOBILE LiDAR SAMPLES

Dataset	Number of points	Length of road section
HDR	1 347 044 219	9922 m
HCR	1 400 438 347	10 822 m
RS-1	127 599 250	800 m
RS-2	78 556 878	570 m
RS-3	63 360 079	400 m

Then, the class label L^* of vector $\boldsymbol{v}_{\text{in}}$ associated with a traffic-sign is determined as follows:

$$L^* = \arg\max_n \left\{v_{\text{out}}^{(n)}\right\}. \tag{11}$$

Before we perform the supervised GB-DBM classifier to classify a traffic sign, we first resize the traffic-sign image into the size of $M \times M$ pixels, and normalize its pixel values into the range [0, 1]. Afterward, the processed image is linearly arranged into a real-valued vector as an input to our supervised GB-DBM classifier, which finally infers the class information of the traffic sign based on (9)–(11).

V. EXPERIMENTS AND DISCUSSION

A. Mobile LiDAR Datasets

This research uses data collected by a RIEGL VMX-450 system. The system is composed of the following: two RIEGL VQ-450 laser scanners, four CCD cameras, a set of Applanix POS LV 520 processing systems containing two global navigation satellite system antennas, an initial measurement unit, and a wheel-mounted distance measurement indicator. The surveyed area is in Xiamen Island, Xiamen, China. A complete survey, at approximately 50 km/h, was conducted along Huandao Road from Xiamen University to the International Conference and Exhibition Center. This is a typical tropical urban environment with high buildings, dense vegetation, and traffic signposts along the surveyed road. Two scanned point cloud data covering eleven- and ten-kilometer-long road sections (see Table I) were selected to evaluate the performance of our proposed method. To clearly demonstrate the experimental results, we selected three small road-section datasets from the two road datasets: RS-1 [see Fig. 6(a)], RS-2 [see Fig. 6(b)], and RS-3 [see Fig. 6(c)] (see Table I).

B. Traffic-Sign Detection

To evaluate the performance of the proposed traffic-sign detection method, we applied it to the above-mentioned mobile LiDAR data and digital images. As seen in Fig. 6, the mobile LiDAR road section samples show typical urban environment with dense shrubs, hedges, and trees, leading to some occlusions and shadows of traffic signs. According to the information for the road sections, we empirically selected the parameters used in our study. The parameters are listed in Table II.

The prior knowledge of the surveyed road suggested that $\Gamma_{\text{RW}} = 12.0\,\text{m}$. The trajectory data record the position

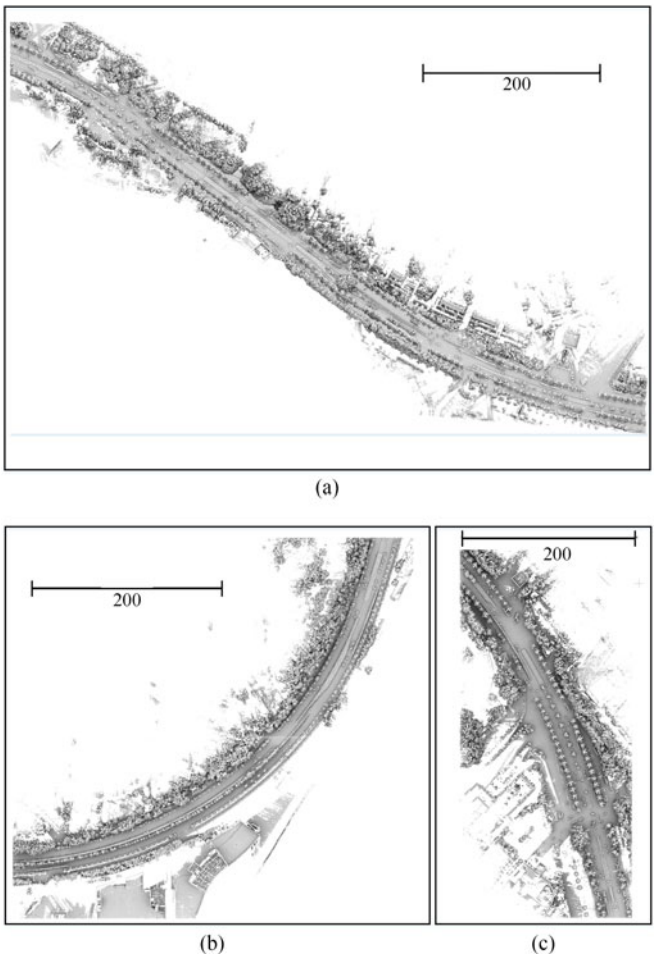


Fig. 6. Point cloud datasets: (a) RS-1, (b) RS-2, and (c) RS-3.

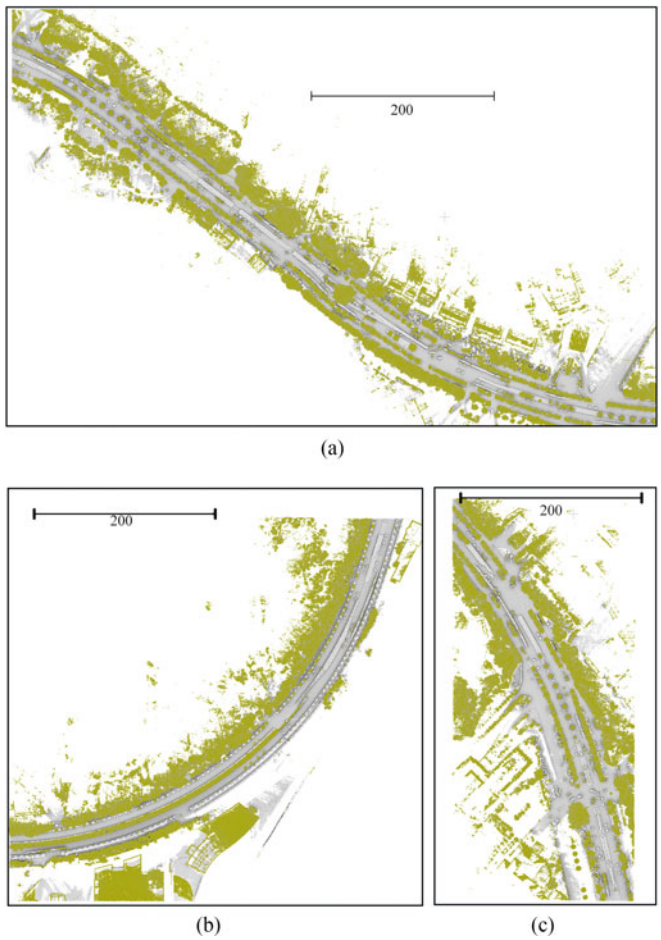


Fig. 7. Filtered results with the threshold of pole-height: (a) RS-1, (b) RS-2, and (c) RS-3.

TABLE II
PARAMETERS USED FOR LOCATING TRAFFIC SIGNS

Parameter	Value
Pole height (A_{PH})	$\Gamma_{\mathrm{PH}} = 1.0$ (m)
Road width (A_{RW})	$\Gamma_{\mathrm{AW}} = 12.0$ (m)
Intensity (A_I)	Γ_I was estimated from the manually selected traffic signs in survey areas
Geometrical structure (A_G)	Planar
Traffic-sign size (A_A)	$\Gamma_A = 0.2$ (m^2)
Voxel size (V_s)	$\Gamma_S = 0.05$ (m)
Clustering distance (C_{dist})	$\Gamma_{\mathrm{dist}} = 0.1$ (m)

information of the moving vehicle. In this survey, the height of the vehicle was 1.6 m. Thus, according to the trajectory data, the value of Γ_{PH} was set to 0.6 m. Using the predefined Γ_{RW} and Γ_{PH}, nonground points were extracted from the road samples (see olive shading in Fig. 7). In this survey, the average point density of the collected points was about 250 points/m^2. Accordingly, the extracted nonground points were voxelized with $\Gamma_S = 0.05$ m.

Moreover, the traffic-sign intensity threshold Γ_I was automatically estimated from dozens of traffic signs in the surveyed data. We calculated the intensity values and covariance matrices for each voxel. The geometrical structure was obtained by analyzing the eigenvalues of the covariance matrix. After traffic-sign labeling, a refinement procedure was performed. A clustering with $\Gamma_{\mathrm{dist}} = 0.1$ m was applied to the labeled traffic-sign voxels. As we mentioned, the traffic-sign size threshold (Γ_A) was empirically set to 0.2 m^2. With the predefined Γ_A, the traffic sign clusters were refined by removing small clusters.

Fig. 8 shows the results of traffic-sign detection (red points). To clearly demonstrate the detected traffic signs, pole-like objects (green points in Fig. 8) were detected by [36]. Visual inspection shows that the detected traffic-sign results were satisfactory and hung on the poles. According to the detected traffic signs, their poles are then determined from clusters in the neighborhoods. The two datasets in this study contain a total of 1268 traffic signs. We detected 1162 traffic signs, including 1101 correctly detected traffic signs and 61 nontraffic signs. The detection accuracy is 86.8%. Some incompletely scanned traffic signs, caused by occlusions, were also undetected because of insufficient salient features. Although some advertising boards attached to light poles were misclassified as traffic signs due to strong reflectance, the majority of traffic signs were correctly detected.

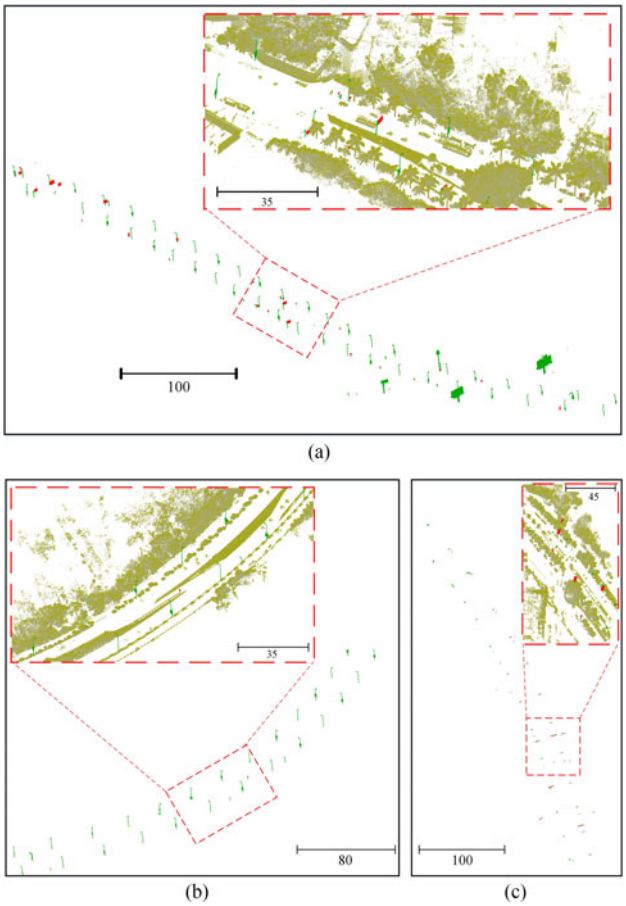


Fig. 8. Traffic sign detection results and their close views: (a) RS-1, (b) RS-2, and (c) RS-3.

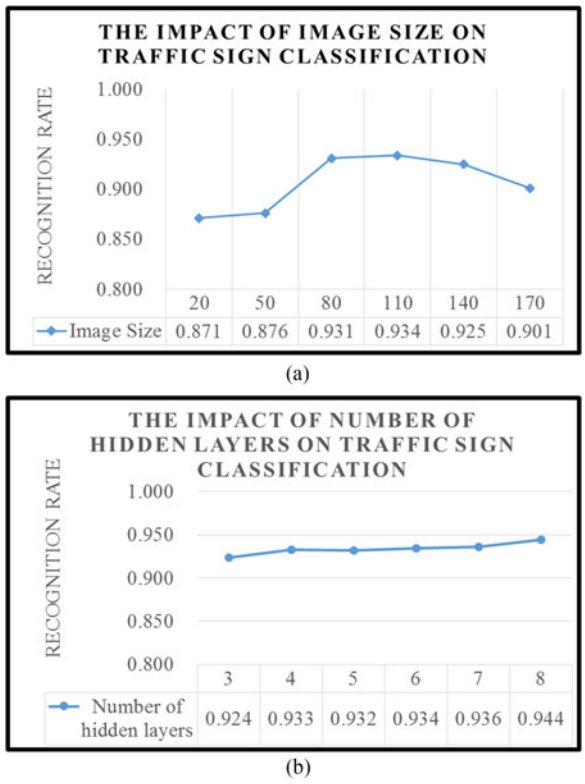


Fig. 9. Parameter sensitivity of (a) pixel size and (b) number of hidden layers.

C. Traffic-Sign Classification

After the traffic-sign interest regions were extracted from the mobile LiDAR points, to obtain their images, we projected them onto the digital images (see Fig. 4). Then, we performed the supervised GB-DBM classifier to classify the resized traffic-sign images into specific categories. In our GB-DBM classifier, parameters—image size (M) and number of hidden layers (N)—were used. The two parameters significantly impact the traffic sign classification results and computational complexity. Thus, we designed two groups of experiments to investigate the sensitivity of our proposed GB-DBM classifier to the selections of N and M.

In this study, we downloaded standard traffic signs from the Ministry of Transport, China, as the training data. Therefore, all the training images and detected traffic-sign images were resized into a size of $M \times M$ pixels. The surveyed road is a coastal landscape road, with several smooth turns, containing roughly 40 traffic-sign categories, according to functionality. The training and test datasets contain 143 360 and 1101 traffic sign images, respectively. We manually labeled the 1101 traffic signs (containing 35 types) of different image conditions as the reference data to evaluate the performance of our GB-DBM classifier.

To quantitatively evaluate the traffic sign classification accuracy, we used the recognition rate, which is defined by the proportion of correctly classified traffic signs.

1) Parameter Sensitivity: In our supervised GB-DBM classifier, two parameters—image size (M) and number of hidden layers (N)—play an important role in traffic-sign classification performance. To obtain the optimal estimation for each of the two parameters and to test the performance of each parameter on the traffic-sign classification results, we conducted two groups of experiments.

In the first group, we held the number of the hidden layers at $N = 3$ and varied image size M from 20 pixels to 170 pixels in intervals of 30 pixels. The test results of the parameters are presented in Fig. 9(a). As shown in Fig. 9(a), the classification performance improves as the image size increases because the larger the image size, the more abundant the image information. However, if the image size exceeds 80, the classification performance changes very slightly. In addition, an increase in the image size causes an increasing computational burden at the training stage. Thus, to balance classification performance and computational burden, we set the image size at $M = 80$ pixels.

In the second group, we kept $M = 80$ pixels and varied the number of hidden layers from 3 to 8 at intervals of 1. The test results of the parameters are presented in Fig. 9(b). As shown in Fig. 9(b), when N changes from 3 to 8, the classification performance smoothly increases because, the larger the number of hidden layers, the more comprehensive and correct the information to be represented. Additionally, an increase of the number of hidden layers leads to increasing computational complexity.

Fig. 10. Illustration of a subset of detected traffic-sign images.

Thus, to obtain promising classification performance, we set the number of hidden layers to 4.

2) Overall Performance: To evaluate the performance of our supervised GB-DBM classifier, we applied it to the mobile LiDAR dataset. After parameter sensitivity analysis, we set $M = 80$ pixels and $N = 4$. In this study, 35 types of traffic signs were tested according to the study scene.

Fig. 10 shows the detected traffic-sign samples on the digital images. We constructed a 6400-1000-1000-500-35 GB-DBM classifier to classify the detected traffic signs into 35 categories. The training time for the supervised GB-DBM classifier was about 4.6 h. For the 1101 test samples, 1027 traffic signs of different shapes and conditions were correctly classified by our supervised GB-DBM classifier, whereas 74 traffic signs were misclassified. Quantitatively, a recognition rate of 93.3% is achieved by our supervised GB-DBM classifier. The misclassification of the traffic signs might be caused by

1) distorted traffic-sign images due to a very large viewpoint;
2) poor traffic-sign image quality due to extremely strong or poor illumination; and
3) incomplete traffic-sign images due to serious occlusion.

3) Computational Performance: Performance validations were performed using an HP pavilion 500-451cn PC with an Intel Core i5-4460 CPU at 3.20 GHz. For point clouds with 1.35 billion (dataset 1) and 1.40 billion (dataset 2) points, the times required by the algorithm to detect traffic signs were around 1.01 and 1.03 h, respectively. To efficiently process the two datasets, we segmented them into road segments with about a road length of 50 m each. Accordingly, a multithread computing environment containing sixteen parallel threads was adopted. By our supervised GB-DBM classifier, the time required to classify the 1101 traffic signs was only about 3.5 min. Due to the use of standard traffic sign databases at the training stage, training results can be used repeatedly in multiple traffic sign classification tasks (one training, multiple uses). Thus, the time required to classify traffic signs is satisfactory for real applications.

4) Comparative Tests: A comparative study was carried out to compare our traffic-sign detection strategy with two newly proposed methods, i.e., Guan's method [36] and Riveiro's method [19]. Guan's method presented correctness, omission, and commission of 88.9%, 11.1%, and 2.8%, respectively, in detecting pole-like road objects using only geometric information. Riveiro's method, which detected traffic signs using geometric and radiometric information, outperformed the other two methods using only geometric information (Pu's method [18] and Yu's method [37] based on pairwise 3-D-shape context). Riveiro *et al.* [19] presented completeness rates of 92.11% for the detection of poles containing traffic signs. Our method achieved a detection accuracy of 86.8%, lower than the newly two methods. However, it achieved a better computing performance in the traffic-sign detection. As we aforementioned, the times acquired for processing the 1.35 billion (dataset 1) and 1.40 billion (dataset 2) points were around 1.01 and 1.03 h, respectively. Riveiro's method reported that the maximum size tested using the system described above was 68 million points. Guan's method presented that the time required for processing 1728 million points was 1.35 h (39 min for generating the contextual visual vocabulary and 42 min for detecting pole-like objects).

In order to further evaluate the performance and feasibility of our supervised GB-DBM method, we compared it with two newly proposed methods—CNN method [13] and Riveiro's method [19]—for classifying traffic signs in the traffic-sign images. The dataset includes 1101 traffic-sign images, containing 35 classes. Riveiro's method presented rates of the efficiency of 83.91% for the classification of traffic signs containing circular, triangular, and rectangular shapes using geometric and intensity information. Due to no use of color information, which remains necessary to efficiently distinguish traffic signs from other pole-like objects, Riveiro's method achieved a relatively low traffic-sign classification accuracy. By means of the CNN method, 1019 traffic signs were correctly detected out of the 1101 test samples, whereas 82 traffic signs were misclassified. Quantitatively, a recognition rate of 92.6% was achieved. Comparatively, our method obtained similar classification accuracies to the CNN method. This is because both the CNN method and our method use high-level feature representations of traffic signs to improve the capability of handling various traffic sign distortions, thereby achieving better traffic sign classification performance.

In conclusion, our proposed two-step traffic-sign detection and classification method is very promising and obtains reliable classification results in mobile LiDAR data. By means of prior knowledge and attributes of traffic signs, traffic-sign candidates can be efficiently detected from very accurate 3-D point clouds. The detected traffic signs are then projected onto digital images and input to our supervised GB-DBM classifier, which exploits high-level feature representations of traffic signs using color information. The traffic-sign detection in 3-D point clouds reduces confusion of similar objects (vehicle license plates, poles, etc.) in road environment, leading to the improvement of traffic-sign classification.

VI. Conclusion

In this paper, we proposed a traffic-sign detection and classification method using mobile LiDAR point clouds and digital images. The traffic-sign detection task was accomplished based on mobile LiDAR point clouds by using a prior knowledge of pole height and road width, intensity, geometrical structure, and traffic-sign size, whereas the classification task was achieved based on digital images by using a supervised GB-DBM model. Our proposed method was evaluated on two datasets collected by a RIEGL VMX-450 system. The detection accuracy of 86.8% was achieved and the classification accuracy of 93.3% was achieved via the supervised GB-DBM classifier. By evaluating computational efficiency and adopting a multithread computing strategy, our proposed method rapidly handles large-volume mobile LiDAR point clouds toward traffic-sign detection and classification. Using mobile LiDAR data, we provide an effective solution to rapid, accurate detection, and classification of traffic signs toward transportation-related applications.

Future research should address the integration of geometric shape, retro-reflectivity, and color information for automatically detecting and classifying traffic signs from large point clouds and digital images collected by mobile LiDAR systems.

Acknowledgment

The authors would like to acknowledge Prof. C. Wang of Fujian Key Laboratory of Sensing and Computing for Smart City, Xiamen University, Xiamen, China, for providing RIEGL VMX-450 mobile LiDAR datasets. The authors would like to thank M. McAllister for his assistance in proofreading the paper. The authors would also like to thank the anonymous reviewers for their valuable comments.

References

[1] A. Gudigar, S. Chokkadi, and R. U, "A review on automatic detection and recognition of traffic sign," *Multimedia Tools Appl.*, vol. 71, pp. 1363–1380, Oct. 2014.

[2] J. M. Lillo-Castellano, I. Mora-Jimenez, C. Figuera-Pozuelo, and J. I. Rojo-Alvarez, "Traffic sign segmentation and classification using statistical learning methods," *Neurocomputing*, vol. 153, pp. 286–299, 2015.

[3] C. Wen *et al.*, "Spatial-related traffic sign inspection for inventory purposes using mobile laser scanning data," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 27–37, Apr. 2016.

[4] M. Brogan, S. McLoughlin, and C. Deegan, "Assessment of stereo camera calibration techniques for a portable mobile mapping system," *IET Comput. Vision*, vol. 7, no. 3, pp. 209–217, Jun. 2013.

[5] X. Yuan, X. Hao, H. Chen, and X. Wei, "Robust traffic sign recognition based on color global and local oriented edge magnitude patterns," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1466–1477, May 2014.

[6] F. Larsson, M. Felsberg, and P. E. Forssén, "Correlating Fourier descriptors of local patches for road sign recognition," *IET Comput. Vision*, vol. 5, no. 4, pp. 244–254, Jun. 2011.

[7] J. Park and K. Kim, "Design of a visual perception model with edge-adaptive Gabor filter and support vector machine for traffic sign detection," *Expert Syst. Appl.*, vol. 40, pp. 3679–3687, 2013.

[8] F. Zaklouta and B. Stanciulescu, "Real-time traffic-sign recognition using tree classifiers," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1507–1517, Nov. 2012.

[9] A. Ruta, Y. Li, and X. Liu, "Robust class similarity measure for traffic sign recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 4, pp. 846–855, Jul. 2010.

[10] S. Xu, "Robust traffic sign shape recognition using geometric matching," *IET Intell. Transp. Syst.*, vol. 3, no. 1, pp. 10–18, Jan. 2009.

[11] D. Pei, F. Sun, and H. Liu, "Supervised low-rank matrix recovery for traffic sign recognition in image sequences," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 241–244, Jan. 2013.

[12] A. Adam and C. Ioannidis, "Automatic road-sign detection and classification based on support vector machines and HOG descriptors," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. II-5, pp. 1–7, 2014.

[13] J. Jin, K. Fu, and C. Zhang, "Traffic sign recognition with hinge loss trained convolutional neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1991–2000, Mar. 2014.

[14] H. Fleyeh and E. Davami, "Eigen-based traffic sign recognition," *IET Intell. Transp. Syst.*, vol. 5, no. 3, pp. 190–196, Jun. 2011.

[15] H. Guan, J. Li, S. Cao, and Y. Yu, "Use of mobile LiDAR in road information inventory: A review," *Int. J. Image Data Fusion*, vol. 7, no. 3, pp. 219–242, 2016.

[16] R. Rybka, "Autodesk and Bentley systems talk about mobile LiDAR," *LiDAR*, vol. 1, no. 2, pp. 41–44, 2011.

[17] K. Williams *et al.*, "Synthesis of transportation applications of mobile LiDAR," *Remote Sens.*, vol. 5, pp. 4652–4692, 2013.

[18] J. Beraldin, F. Blais, and U. Lohr, "Laser scanning technology," in *Airborne and Terrestrial Laser Scanning*, G. Vosselman and H. Mass, Eds. Dunbeath, U.K.: Whittles Publishing, 2010, pp. 1–42.

[19] B. Riveiro *et al.*, "Automatic segmentation and shape-based classification of retro-reflective traffic signs from mobile LiDAR data," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 1, pp. 295–303, Jan. 2016.

[20] Y. Yu, J. Li, C. Wen, H. Guan, H. Luo, and C. Wang, "Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data," *ISPRS J. Photogramm. Remote Sens.*, vol. 113, pp. 106–123, 2016.

[21] Y. Chen, H. Zhao, and R. Shibasaki, "A mobile system combining laser scanners and cameras for urban spatial objects extraction," in *Proc. IEEE Conf. Mach. Learn. Cybern.*, Pokfulam, Hong Kong, vol. 3, 2007, pp. 1729–1733.

[22] H. Yokoyama, H. Date, S. Kanai, and H. Takeda, "Pole-like objects recognition from mobile laser scanning data using smoothing and principal component analysis," *ISPRS, Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 38-5/W12, pp. 115–120, 2011.

[23] 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.

[24] H. Wang *et al.*, "Object detection in terrestrial laser scanning point clouds based on Hough forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 807–1811, Mar. 2014.

[25] S. Pu, M. Rutzinger, G. Vosselman, and S. O. Elberink, "Recognizing basic structures from mobile laser scanning data," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 6, pp. S28–S29, 2011.

[26] C. Ai and Y. J. Tsai, "Critical assessment of an enhanced traffic sign detection method using mobile LiDAR and INS technologies," *J. Transp. Eng.*, vol. 141, no. 5, pp. 1–12, 2015.

[27] X. Chen *et al.*, "Next generation map making: Geo-referenced ground-level LiDAR point clouds for automatic retro-reflective road feature extraction," in *Proc. ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, Seattle, WA, USA, 2009, pp. 488–491.

[28] A. Vu, Q. Yang, J. A. Farrell, and M. Barth, "Traffic sign detection, state estimation, and identification using onboard sensors," in *Proc. Int. IEEE Annu. Conf. Intell. Transp. Syst.*, The Hague, The Netherlands, 2013, pp. 875–880.

[29] I. Puente *et al.*, "Review of mobile mapping and surveying technologies," *Measurement*, vol. 46, pp. 2127–2145, Mar. 2013.

[30] J. Han *et al.*, "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2015.

[31] R. Salakhutdinov, J. B. Tenenbaum, and A. Torralba, "Learning with hierarchical-deep models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1958–1971, Dec. 2012.

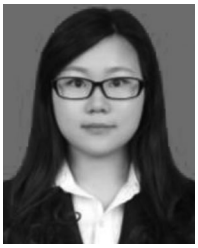
[32] B. Leng, X. Zhang, M. Yao, and Z. Xiong, "A 3D model recognition mechanism based on deep Boltzmann machines," *Neurocomputing*, vol. 151, pp. 593–602, 2015.

[33] Y. Huang, W. Wang, L. Wang, and T. Tan, "Conditional high-order Boltzmann machines for supervised relation learning," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4297–4310, Sept. 2017.

[34] J. Demantké, C. Mallet, N. David, and B. Vallet, "Dimensionality based scale selection in 3D Lidar point clouds," *ISPRS, Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 38, no. W12, pp. 97–102, 2011.
[35] R. Salakhutdinov and G. Hinton, "An efficient learning procedure for deep Boltzmann machines," *Neural Comput.*, vol. 24, no. 8, pp. 1967–2006, 2012.
[36] H. Guan *et al.*, "Pole-like road object detection in mobile LiDAR data via supervoxel and bag-of-contextual-visual-words representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 4, pp. 520–524, Apr. 2016.
[37] Y. Yu *et al.*, "Semiautomated extraction of street light poles from mobile LiDAR point-clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1374–1386, Mar. 2015.
[38] S. Pu *et al.*, "Recognizing basic structure from moble laser scanning data for road inventory studies," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 6, pp. 28–39, 2011.

Haiyan Guan (M'15) received the Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009, and in geomatics from the University of Waterloo, Waterloo, ON, Canada, in 2014.

She is currently a Professor with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing, China. She has authored or co-authored more than 40 research papers in refereed journals, books, and proceedings, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *ISPRS Journal of Photogrammetry and Remote Sensing*, and IGARSS and ISPRS proceedings. Her current research interests include information extraction from LiDAR point clouds and from earth observation images.



Wanqian Yan received the Bachelor's degree in remote sensing science and technology in 2017 from Nanjing University of Information Science and Technology, Nanjing, China, where she is currently working toward the Master's degree at the School of Remote Sensing and Geomatics Engineering.

Her current research interests include information extraction from LiDAR point clouds and machine learning.

Yongtao Yu (M'16) received the Ph.D. degree in computer science and technology from Xiamen University, Xiamen, China, in 2015.

He is currently an Assistant Professor with the Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian, China. He has co-authored more than 30 research papers in refereed journals, books, and proceedings. His research interests include pattern recognition, machine learning, intelligent interpretation of 3-D point clouds, and remotely sensed imagery.

Liang Zhong received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2010.

He is currently a Senior Engineer with Changjiang Spatial Information Technology Engineering Co., Ltd., Wuhan, China. He has authored or co-authored more than 15 research papers in refereed journals, books, and proceeding, including the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and *ISPRS International Journal of Geo-Information*. His current research interests include telemetry processing and LiDAR data processing.

Dilong Li received the Master's degree in computer science and technology from Xiamen University, Xiamen, China, in 2014. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing at Wuhan University, Wuhan, China.

His current research interests include pattern recognition, machine learning, information extraction from LiDAR point clouds, and remotely sensed imagery.