ISPRS Journal of Photogrammetry and Remote Sensing 160 (2020) 149–166

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Automated extraction of lane markings from mobile LiDAR point clouds based on fuzzy inference

Heidar Rastiveis[a,*], Alireza Shams[b], Wayne A. Sarasua[c], Jonathan Li[d,e]

[a] Department of Photogrammetry and Remote Sensing, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran
[b] Advanced Highway Maintenance & Construction Technology (AHMCT) Research Center, University of California, Davis, CA, USA
[c] Glenn Department of Civil Engineering, Clemson University, Clemson, SC, USA
[d] Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, China
[e] Departments of Geography and Environmental Management & Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

ARTICLE INFO

Keywords:
Mobile LiDAR
Road lane markings
Point cloud
Fuzzy inference system

ABSTRACT

Mobile LiDAR systems (MLS) are rapid and accurate technologies for acquiring three-dimensional (3D) point clouds that can be used to generate 3D models of road environments. Because manual extraction of desirable features such as road traffic signs, trees, and pavement markings from these point clouds is tedious and time-consuming, automatic information extraction of these objects is desirable. This paper proposes a novel automatic method to extract pavement lane markings (LMs) using point attributes associated with the MLS point cloud based on fuzzy inference. The proposed method begins with dividing the MLS point cloud into a number of small sections (e.g. tiles) along the route. After initial filtering of non-ground points, each section is vertically aligned. Next, a number of candidate LM areas are detected using a Hough Transform (HT) algorithm and considering a buffer area around each line. The points inside each area are divided into "probable-LM" and "non-LM" clusters. After extracting geometric and radiometric descriptors for the "probable-LM" clusters and analyzing them in a fuzzy inference system, true-LM clusters are eventually detected. Finally, the extracted points are enhanced and transformed back to their original position. The efficiency of the method was tested on two different point cloud datasets along 15.6 km and 9.5 km roadway corridors. Comparing the LMs extracted using the algorithm with the manually extracted LMs, 88% of the LM lines were successfully extracted in both datasets.

1. Introduction

Within the last two decades, airborne Light Detection and Ranging (LiDAR) scanning technologies have been widely used to acquire three-dimensional (3D) coordinates in the form of dense point clouds for various applications such as 3D city modeling and Digital Elevation Model (DEM) generation (Guan et al., 2014; Liu, 2008; Yang et al., 2017). Similarly, Mobile Laser Scanning (MLS) systems provide an efficient, cost-effective, and reliable data source in the form of 3D point clouds for corridors, which is useful in roadway analysis (Guan et al., 2014; Hatger and Brenner, 2003; Shams et al., 2018; Wang et al., 2017; Yan et al., 2016; Zai et al., 2017). An MLS includes an integration of several devices including a laser scanner, a Global Navigation Satellite System (GNSS), an Inertial Measurement Unit (IMU), high-resolution cameras, and a computer control device (Kumar et al., 2017).

Due to the tremendous number of the points within the resultant point cloud, the manual extraction of useful data is tedious; therefore, researchers have been working on developing automatic techniques for filtering, segmentation, and classification of point clouds to extract desirable objects (Habib et al., 2005; Jiang, 2017; Ma et al., 2018; Serna et al., 2014; Wen et al., 2019; Zai et al., 2017).

This paper proposes a novel automatic method to extract pavement LMs using point attributes associated with the MLS point cloud based on fuzzy inference. A potential benefit of using fuzzy inference is that it may address limitations of other methods that are prone to uncertainty. One of the objectives of this research is to explore the benefits of using fuzzy inference in dealing with the uncertainty associated with extracting roadway markings automatically.

2. Literature review

2.1. Extraction of street features from LiDAR point clouds

There have been numerous studies focused on extracting street

* Corresponding author.
E-mail addresses: hrasti@ut.ac.ir (H. Rastiveis), asham@ucdavis.edu (A. Shams), sarasua@clemson.edu (W.A. Sarasua), junli@xmu.edu.cn, junli@uwaterloo.ca (J. Li).

https://doi.org/10.1016/j.isprsjprs.2019.12.009
Received 3 June 2019; Received in revised form 23 November 2019; Accepted 10 December 2019

features and roadway assets from point clouds. For example Wang et al. (2017) proposed a feature matching algorithm for automatic street object recognition in an MLS point cloud. Their algorithm used shape descriptors of complete objects to match repetitive objects in large point clouds to identify and extract traffic signs, lamp poles, roadside furniture, and other objects. Many studies have investigated the extraction of curbs and road surfaces by using differences in height or projected density thresholds (Jaakkola et al., 2008; Guan et al., 2014; Rodríguez-Cuenca et al., 2015; Yang and Dong, 2013). Further, the point clouds can be used to extract roadway geometry elements such as horizontal alignment (Holgado-Barco et al., 2015) and cross sections including cross slope and super-elevation data that is relatively accurate for highway safety applications (Holgado-Barco et al., 2017; Shams et al., 2018).

The intensity of point data facilitates the automatic extraction of pavement markings from MLS point clouds in all of the studies identified in the literature (Guan et al., 2014; Kumar et al., 2014; Shams et al., 2018; Yang et al., 2012). Olsen et al. (2018) investigated the effectiveness of using MLS to evaluate pavement marking and highway sign retro-reflectivity according to the intensity value of the points. Moreover, Ai and Tsai (2016) developed an automated system to determine the condition of traffic signs using MLS. One benefit of using MLS to determine retro-reflectivity of pavement marking and sign is that the data can be collected during the daytime and nighttime.

2.2. Methods of extraction of pavement markings from MLS point clouds

Based on the literature, LM extraction algorithms can be divided into two groups: (1) points-based; and (2) raster-based. These methods are discussed in the following sub-sections.

2.2.1. Points-based methods

In the points-based methods, an algorithm directly applies the raw points and their available information to extract road markings. For instance, Vosselman (2009) extracted road markings points by a range of dependent thresholds and grouped them by connected component analysis. The road markings were extracted by fitting predefined shapes to the grouped segments (Yang and Dong, 2013) used support vector machine (SVM) classification techniques to classify point clouds according to calculated geometric features of each point in a specified neighborhood. Soilán et al. (2017) applied a curb-based approach to segment the raw point cloud, and then directly extract road markings from road surface points through multi-segment thresholding and spatial density filtering. Yang et al. (2017) applied a binary kernel descriptor (BKD) consisting of Gaussian kernel density estimation and binarized components to encode the shape and intensity of the 3D point clouds to extract road information from MLS point clouds. This information was used in a random forest classifier to extract curbs and markings on the road with a 94.5% level of accuracy. Zai et al. (2018) used geometric, intensity and spatial distance information to generate super-voxels from point clouds, and then 3D road boundaries were extracted using the α-shape algorithm and the graph cuts-based energy minimization algorithm. In this method, average completeness over 95% and average correctness over 98% were reported.

2.2.2. Raster-based methods

In raster-based methods, the point's intensity and range data are converted into a raster format to create a 2D image, and image processing algorithms are used to extract road markings. For example, Jaakkola et al. (2008) corrected LiDAR intensity data using a second order curve fitting function and then applied threshold and morphological filtering to extract road markings. Other variations of raster-based methods to extract road markings were used by Chen et al. (2009), Smadja et al. (2010), Yang et al. (2012), Kumar et al. (2014), and Guan et al. (2015). Soilán et al. (2018) applied Artificial Neural Network (ANN) for extraction and classification of pavement markings based on Geometry Based Feature (GBF) vectors including area, bounding ellipse measurements, shape measurement, and pixel distribution. An average accuracy of 92% for three different datasets was reported for this algorithm.

A state-of-the art raster-based method was recently proposed in Jung et al. (2019) to extract road pavement markings from MLS point clouds. In this method, using the MLS trajectory information, the point cloud was divided into smaller sections. Then, the road surface was extracted via constrained RANSAC and the segments were converted to a 2D intensity image to apply image segmentation to separate the LMs from the road pavement. The algorithm was tested on a variety of datasets and F1-score measures from 89 to 97% were reported for a wide range of road scene types. Moreover, Wen et al. (2019) developed a deep learning-based framework for automated extraction and classification of road markings. In their technique, a pixel-level U-net segmentation network was trained using a series of raster patches from the road surface. The precision, recall, and F1-score obtained using the low-quality point cloud datasets achieved 95.97%, 87.52% and 91.55%, respectively.

2.3. Proposing a combined approach

Generally, converting the points into a raster format enables using numerous image processing techniques which can be highly efficient in terms of simplicity and speed of processing. However, raster conversion may result in a loss of information. In contrast, working with points directly prevents this drawback but requires complex computation. In this paper, we propose a new method by taking advantage of the strengths of both point-based and raster-based methods to extract LMs. We first use a raster format to estimate the primary location of LMs to increase the speed of computation. Then, the points are extracted and processed to detect LMs.

To date, intensity of the points is the only attribute used to extract LMs in most extraction methods. Other attributes such as trajectory data, Number of Returns (NOR), and Return Number (RN) can also be useful. NOR shows the maximum number of individual returns (up to five) that can be extracted from a single beam. RN refers to the rank of a return among those generated from one beam and it is meaningful only in a system that supports multiple returns per beam (Gatziolis and Andersen, 2008). As part of our research we explored the benefits of using RN information to detect LMs. This is one of the unique aspects of this research based on the literature review. Moreover, similar to many MLS data processing studies it is assumed that vehicle trajectory is available. While MLS vehicle trajectory data may not be provided to the end user, it is always initially collected via GNSS/IMU.

In our approach, LMs within each tile are assumed to be straight lines because the curvature of the road is minimal if the tile size is small. Fusing the linearity assumption with the intensity information leads to defining a new descriptor that we call Lane Marking Index (LMI). However, using an intensity threshold as the only attribute to identify pavement markings may return erroneous results depending on the condition of the markings. On the other hand, high-intensity values of the reflected lasers may come from other objects such as adjacent vehicles in the traffic stream or guardrail. Thus, there is clearly uncertainty in a threshold based extraction process. One approach to minimize uncertainty is to use a fuzzy inference system that is designed to analyze and address this uncertainty. As stated previously, one objective of this research is to design a fuzzy inference system that addresses thresholding uncertainty. In the next section, we identify and discuss the novel workflow methodology used for automatically extracting pavement markings. This is followed by a discussion of experimental results, and conclusions.

3. Method

The workflow of the proposed method is shown in Fig. 1. According

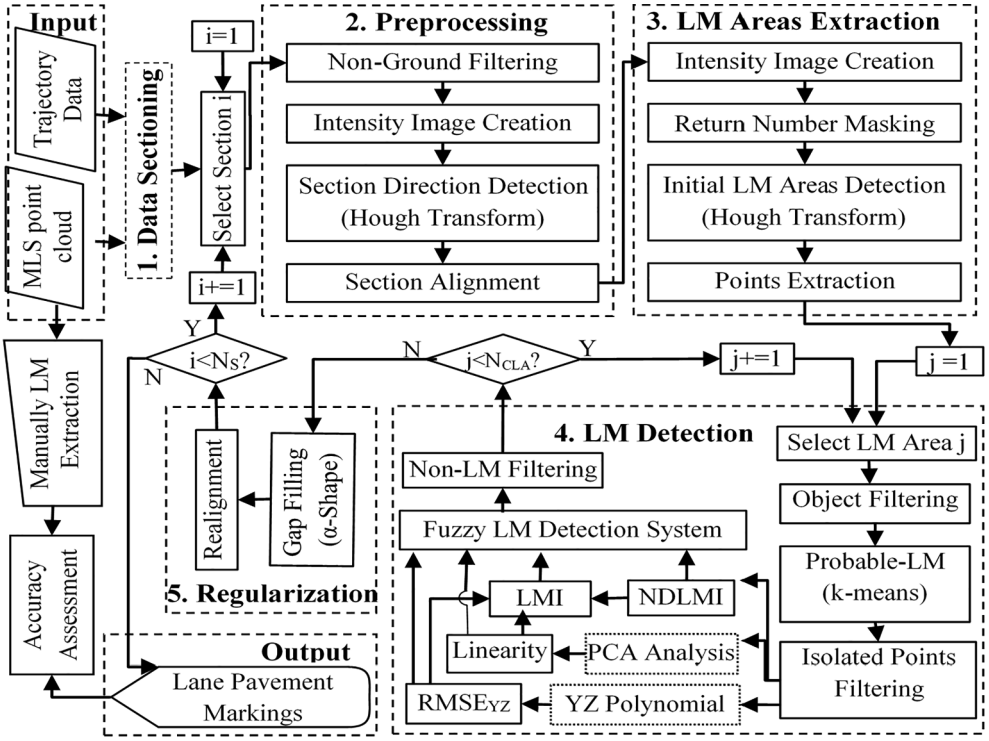

Fig. 1. Flowchart of the proposed method for automatic LM points extraction from MLS point clouds.

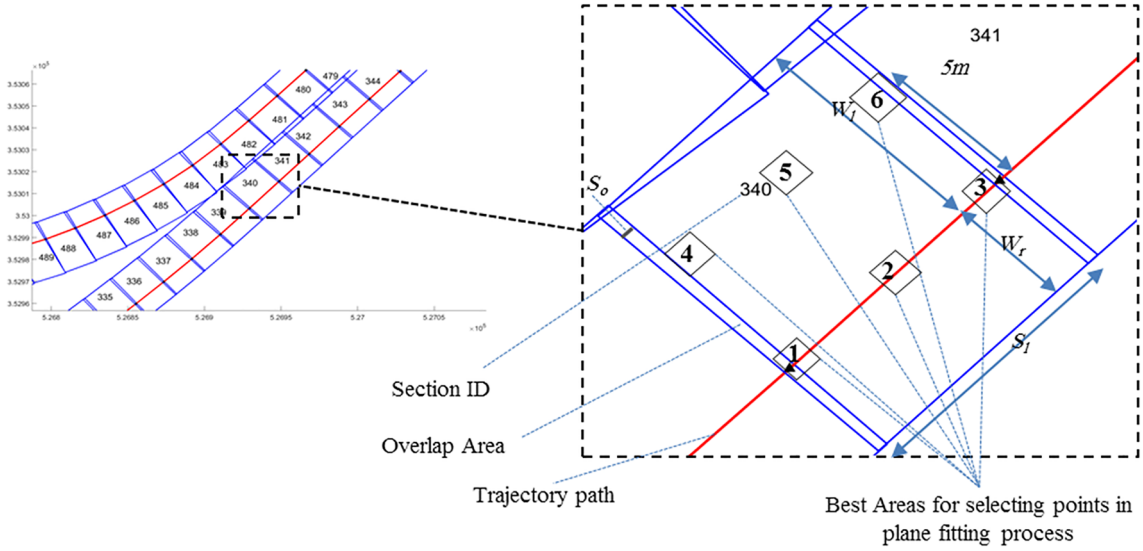

Fig. 2. Sample section areas based on the trajectory data of the vehicle.

to this workflow, pavement markings are extracted from the point cloud and trajectory data in five main phases: (1) data sectioning, (2) preprocessing, (3) lane marking areas extraction, (4) lane markings detection, and (5) regularization.

In this workflow, after dividing the MLS point cloud into small sections (e.g., tiles) along the route, each section is vertically aligned. The initial traveling LM areas are detected using the HT algorithm and considers a buffer area where all points outside the road are omitted from the selected section. In each LM area, candidate pavement marking points are separated from other points. Then, by analyzing their geometric and intensity descriptors in a fuzzy inference system, LMs of each section are detected. Eventually, the points in each LM are refined and transformed back to their original position in the regularization step. Further details of these phases are described in the following sections.

3.1. Data sectioning

An MLS point cloud contains numerous objects with a variety of shapes and sizes. The sectioning step will result in faster processing by dividing the data into a number of small tiles and reducing redundant points from the point cloud. Section length is an important parameter to fulfill the shape of an LM in each section–especially sections of a small radius curve. Although a fixed length in the sectioning step is simple, it may result in a curved LM assumed to be a straight line depending on section length. On the other hand, a tangent section of the road could be over-segmented. Therefore, the size of the tiles should be varied depending on roadway geometry and curve radii.

Varying tile size based on roadway curvature can be done automatically in some instances if MLS vehicle trajectory data is available. For curved sections, smaller tiles can be specified as a function of curve radius that can be easily extracted from the vehicle trajectory. Road curvature may cause gaps between adjacent sections due to the difference between their azimuths. To account for these gaps, overlapping of two adjacent tiles is necessary to prevent gaps during sectioning. The overlap distance can be user specified based on curve radii. A fixed overlap distance of 1 m was used in this research and was found to be suitable for most of the sections.

If vehicle MLS trajectory data is not available, a manual polyline that replicates the trajectory can be drawn. Fig. 2 shows the corresponding parameters in sectioning, where S_l is the section length, and S_o is the overlapping distance between two adjacent sections. Although the sectioning step reduces the volume of data, many redundant points remain in each tile. In this regard, buffering distances across the trajectory line as shown by W_l and W_r in Fig. 2, can help to eliminate undesirable points. If the road width changes along the vehicle path, this parameter can be manually changed. However, for automatic processing during this step one can consider the maximum road width to define a fixed S_w. After finding the section area based on the abovementioned parameters, the overlap area can easily be added to the resulting boundary along the S_l and based on the S_o.

3.2. Preprocessing

The preprocessing phase includes non-ground filtering and section alignment steps which help to simplify the LM extraction process. In the non-ground filtering step, tree branches, poles, utility cables, traffic signs, and bridges above the road surface are eliminated. A variety of research for ground surface extraction from point clouds is available in the literature such as a RANSAC-based method proposed by Jung et al. (2019). While these approaches are effective, they were deemed inefficient and unnecessary for the level of filtering needed in the preprocessing step. Thus, a fast and easy-to-implement approach is proposed for removing objects that uses a two-dimensional plane fitted to a number of selected points on the road surface. Any points whose distance to the fitted plane beyond predefined thresholds are considered as non-ground points.

To fit a surface to the points, first six 0.5 × 0.5 sq.m. areas are selected according to Fig. 2: areas 1, 2 and 3 are from the start, middle, and end of trajectory line, respectively. Similar areas are selected on a parallel line with a 5 m offset to the left (areas 4, 5, and 6). For areas on the trajectory line (areas 1, 2, and 3) the road surface height (H_{rs}) can be obtained from trajectory elevation if it is provided. This assumes that the trajectory elevation has been adjusted for the height of the GNSS antenna above the road surface. Alternatively, if the trajectory does not have accurate elevation data, sampled points from the point cloud are used to determine the H_{rs} for areas 1, 2, and 3. To do this, the sampled points within each area are sorted and divided into four groups with similar elevations, then the median height of the 2nd lowest group is selected as the H_{rs}. The two highest groups will usually include assets above the roadway such as overpasses, and the lowest group is removed due to its potential to be noisy points. Since the road surface on the trajectory path will always have the highest density of points it is expected that the 2nd lowest group or, at least, the majority of its points represent the points on the road surface. A similar procedure is used for areas 4, 5, and 6, however only groups of points within ± 1 m of the H_{rs} points of area 1, 2, and 3 are considered. Once the H_{rs} heights are obtained in all six areas, the points in each area whose heights are similar to the corresponding H_{rs} are extracted and selected for plane fitting.

After filtering the non-ground points each candidate section is aligned either in the X or Y direction to reduce the complexity of computation. The main direction of the candidate section is detected using the HT algorithm, and then a rotation transformation is performed. To that end, first, a low-resolution intensity image of each section is generated and edited using the "*Closing*" morphological operator based on a disk-shaped structural element (SE) with a 2 pixels radius. All lines on the image are extracted by first extracting edge pixels of the image and then performing the HT algorithm. To avoid extracting inconvenient lines as well as speed up the process a specified angle range based on the azimuth of the trajectory data of the candidate section is considered. Thus, if the azimuth of the trajectory line is α, then the angle intervals that can be considered are $[\alpha\text{-}\theta, \alpha+\theta]$. After extracting all lines, the most frequent rotation of the extracted lines is deliberated as the main direction of that section. Eventually, by generating a 2D rotation matrix based on this direction, all points are transformed into a new position.

3.3. Lane marking areas extraction

In this step, a higher resolution intensity image is initially created using the rotated points. Then, the possible vegetation areas are eliminated by applying an RN mask on the obtained image. The RN usually varies from 1 to 6 and can make a bright differentiation between vegetation area and other areas in a point cloud (Gatziolis and Andersen, 2008). To generate the RN mask, each pixel that includes points with an RN value equal to or larger than 2 may be labeled as a vegetation area. The generated RN mask is modified using morphological operators *Dilation* and *Erosion* to fill the gaps and delete noisy pixels considering line-shaped SE shown by Eqs. (1) and (2). In these equations, v is half of the section length and h is the minimum distance to a vegetation area from the road surface.

$$SE = [1 \;\; 1 \;\; \cdots \;\; 1 \;\; 1]^{T}_{1\times v} \tag{1}$$

$$SE = [1 \;\; 1 \;\; \cdots \;\; 1 \;\; 1]_{1\times h} \tag{2}$$

The resulting masked image with the RN mask is further modified using the *Closing* operator considering a disk-shaped SE with a 2 pixels radius. Then, the *Sobel* edge extraction operator can extract the edges along the edited image section. Once the edges are detected, the HT algorithm is executed to extract proper lines that may help to detect the LM areas. Due to the alignment of the section, a small angle interval $[\text{-}\theta, +\theta]$ in the HT can be considered to reduce the computation time. Finally, by considering a buffer space around each line larger than the standard LM width, the LM area can be obtained, and then all points inside this area are extracted. These points are then analyzed in the next step to find the LM points.

3.4. Lane markings detection

After filtering the object points in each LM area, a group of "probable-LM" cluster points is separated. After extracting descriptors from the point cloud, the true-LM points are detected based on a designed fuzzy lane marking detection system. These steps are discussed in the following subsections.

3.4.1. Probable-lane markings extraction

In this step, cars and trees related points that might not have been deleted during the preprocessing step are initially filtered. For this purpose, the minimum elevation in the neighborhood of each point is considered as the assumed ground surface. Then, all points with height differences from this elevation greater than a threshold will be removed. This technique is similar to applying the Normalized Digital Surface Model (nDSM) to point clouds to generate DTMs, which have been used in number of LiDAR point cloud studies (Bartels and Wei, 2010; Höfle et al., 2009; Rastiveis, 2015).

After eliminating noisy points, all remaining points will fall in "probable-LM" and "non-LM" groups. Due to difference in intensity, the "probable-LM" group can be differentiated from the "non-LM" group by various thresholding methods such as Otsu's algorithm, Entropic methods, or Minimum Error methods (Morse, 2000; Sahoo et al., 1988).

However, due to its high speed of computation in addition to its performance, the *k*-means algorithm is applied. The main drawback of this algorithm is the possibility of converging to a local minimum based on poor initial cluster centroids (Jain, 2010). To overcome this weakness, the best result may be selected after running the algorithm a few times and comparing the final cost function. The Sum of the Squared Errors (SSE) over two clusters of "non-LM" and "probable-LM" is calculated as a cost function that is shown by Eq. (3).

$$SSE(C) = \sum_{i=1}^{N_{NL}} \|I_i^{NL} - \mu_{NL}\|^2 + \sum_{j=1}^{N_{PL}} \|I_j^{PL} - \mu_{PL}\|^2 \quad (3)$$

In each "probable-LM" cluster, a number of isolated points may appear which are eliminated based on a density-based analysis. Here, instead of using a cube- or sphere- shape when applying density-based filtering algorithms, a narrow rectangular-shape along the lane direction (Y axis) with a small thickness across the Z axis may be more appropriate. Otherwise, some of the probable-LM points, specifically on damaged LMs, may be mistakenly filtered. The minimum number of points in the neighboring window can be assigned according to the resolution of the LiDAR scanner.

3.4.2. Descriptor extraction

Although the intensity of the remaining points is higher than the eliminated cluster, there is a possibility that these points are not LMs. To minimize the possibility of non-LMs being classified as true-LMs, the remaining points should be checked further. In this study, the true-LMs are detected by extracting a number of geometric and radiometric descriptors for each "probable-LM" cluster and analyzing them in a fuzzy inference decision-making system. A perfect descriptor indicates a clear difference between the true-LM cluster and non-LM cluster. The very first descriptor for detecting true-LMs can be obtained based on the contrast between the intensity of the "probable-LM" clusters and its background or the "non-LM" clusters. Here, the Normalized Difference Lane Marking Index (*NDLMI*) computed by Eq. (4) is defined to measure this contrast.

$$\text{NDLMI} = \frac{\mu_{PLM} - \mu_{NLM}}{\mu_{PLM} + \mu_{NLM}} \quad (4)$$

This measure is higher for the true-LM clusters that have a greater intensity in comparison to the asphalt points, and is lower for false-LM clusters. However, the intensity of damaged LM points may be lower than expected values. Further, LM retro-reflectivity degrades over time and their intensity values may vary greatly as the LM age and reflective beads are dislodged (Sarasua et al., 2003; Thamizharasan et al., 2003). Thus, there is not a consistent boundary or threshold between the *NDLMI* of the true-LM clusters and false ones. This fact proves that not only is there uncertainty in using this *NDLMI* but also necessitates the use of other descriptors as well.

In addition to the *NDLMI* obtained based on the intensity information, geometric properties of traffic pavement LMs may be useful. Here, *Linearity* and $RMSE_{yz}$ values are defined and applied as geometric descriptors for detecting true-LM. *Linearity* can be measured using Principal Component Analysis (PCA). PCA can be defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance lies on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on (Jolliffe, 2011). The principal components of points *P* are the eigenvectors of $C_P = \frac{1}{n}\vec{P}\vec{P}^T$, which $\vec{P}$ is the vector of points after subtracting centroid coordinates. After computing the eigenvalues and eigenvectors of C_P, where eigenvalues are sorted as $\lambda_1 > \lambda_2 > \lambda_3$, the *Linearity* (*L*) value for the points is computed using normalized eigenvalues ($\bar{\lambda}$) from Eq. (5):

$$Linearity = \sqrt{\frac{\bar{\lambda}_1}{\bar{\lambda}_2 + \bar{\lambda}_3}} \quad (5)$$

A higher linear distribution of points results in a bigger first eigenvalue, and consequently greater *Linearity*. Therefore, in comparison to a false-LM cluster, a higher value of *Linearity* for the true-LM cluster is expected. Nonetheless, similar to the *NDLMI* measure there is not a clear threshold for this measure.

Due to the short length of the sections, another geometric property can be defined based on the assumption that the LM points are located on a plane or, more accurately, a parabolic surface with small variation in height information. Assuming a 3D coordinate system that the Y and X axes are in the direct of section length and width respectively, each point group may be fitted to a 2D polynomial in the YZ direction. The Root Mean Square Error of the points ($RMSE_{yz}$) shown by Eq. (6) is used to determine fitness of the polynomial in LM detection phase.

$$RMSE_{yz} = \sqrt{\frac{\sum_{i=1}^{N}(Z - \tilde{Z})^2}{N}} \quad (6)$$

In this equation, *Z* and $\tilde{Z}$ are the original height and the calculated height using the fitted 2D polynomial respectively, and *N* is the number of points in each "probable-LM" cluster. This measure has a small value for true-LM clusters, and a higher value for false-LM clusters such as guardrail or possible passing vehicles.

To apply the abovementioned descriptors in detecting true-LM, one can define a Lane Markings Index (*LMI*) as shown by Eq. (7) to make a distinct boundary between true- and false-LM. Because of the higher value of *NDLMI* and *Linearity*, and the lower value of the $RMSE_{yz}$ a high value of this index is expected in the case of true-LM.

$$LMI = \frac{NDLMI.\ Linearity}{RMSE_{yz}} \quad (7)$$

Although this index may be a powerful descriptor to find true-LMs, because of the uncertainty or fuzziness of the values used to calculate *LMI*, it may not be robust enough to successfully identify all LMs. Thus, this measure along with other descriptors are applied in a fuzzy rule-base inference system to detect true-LMs.

3.4.3. Fuzzy lane marking detection system

Fuzzy theory introduced by Zadeh (1978), which is similar to human reasoning to approximate information to generate decisions, is a useful tool in dealing with problems involving uncertainty (Cox et al., 1998; Zimmerman, 1996). In this research, the Mamdani fuzzy rule-based system is introduced to analyze the extracted features to detect true-LMs. Any fuzzy rule base system is comprised of three main parts of fuzzification, inference, and defuzzification (Rastiveis et al., 2013). The overall view of the designed fuzzy system is illustrated in Fig. 3.

In the first step, fuzzification, the input/output features are divided into fuzzy spaces, specified by fuzzy membership functions that maps the inputs onto the range of [0, 1]. Monotonic, triangular, trapezoidal, and bell-shaped functions are the most frequently used in fuzzy rule based experiments (Mather and Tso, 2016). In this LM detection algorithm, extracted measures (*NDLMI*, $RMSE_{yz}$, *Linearity* and *LMI*) are applied as input variables and the *LM-Probability* is considered as the

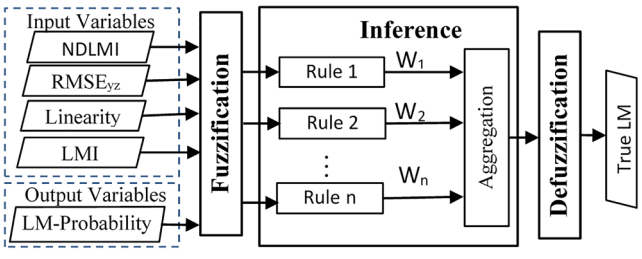

Fig. 3. Overall view of the proposed rule-based fuzzy LM detection system.

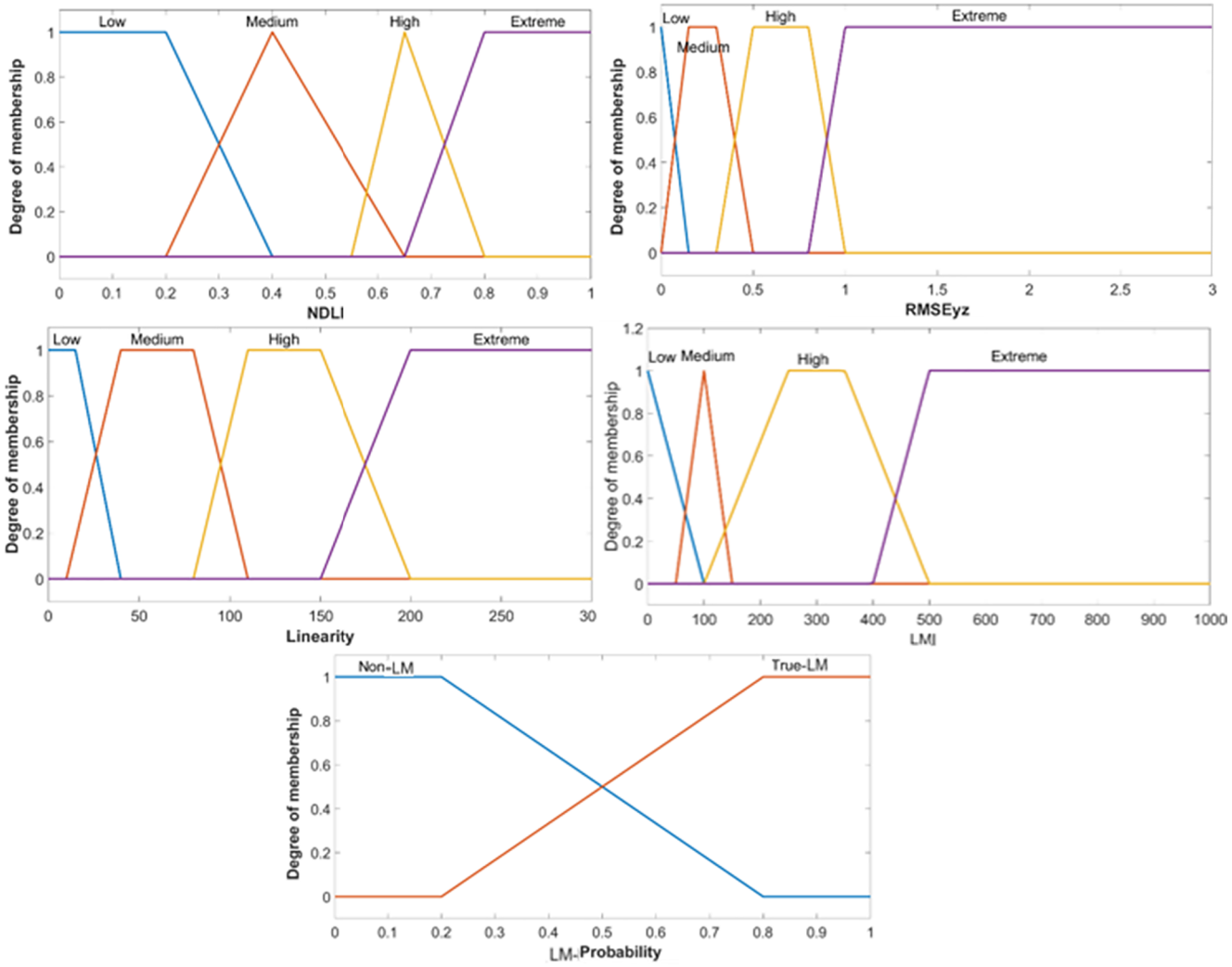

Fig. 4. Assigned membership functions for the linguistic labels of the input and output linguistic variables.

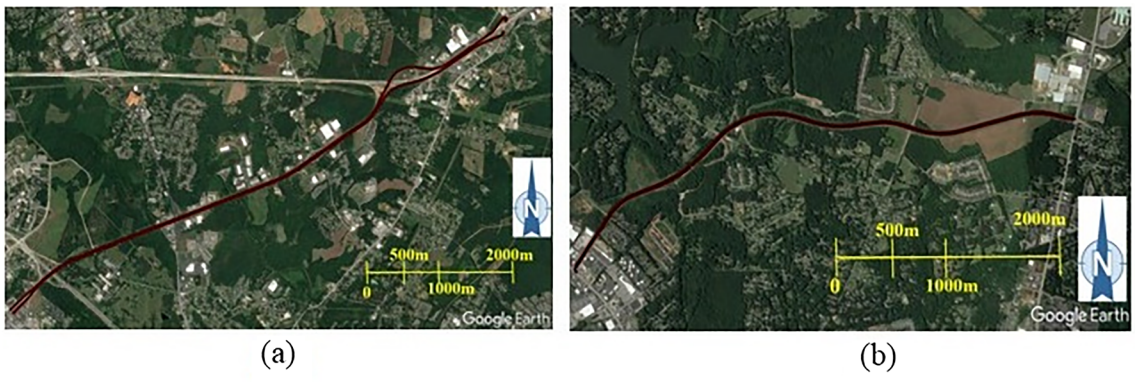

Fig. 5. Overview of the examined datasets: (a) the I-85 BL dataset about 15.6 km; (b) the EWPkwy dataset about 9.5 km.

output variable. Considering trapezoidal- and triangular-shaped functions, an expert has accurately defined them. Fig. 4 illustrates the designed membership functions and linguistic labels for both input and output variables. In this figure, the range of *NDLMI and LM-Probability* limited to [0, 1], but $RMSE_{yz}$, *Linearity*, and *LMI* can be greater than the displayed range. In other words, the upper bound of "*Extreme*" linguistic label in these features is limitless.

A number of IF-THEN rules must be defined in order to import user knowledge in the fuzzy reasoning system. A single fuzzy IF–THEN rule can be formulated according to: *IF x is A; THEN y is B*. In this statement, *A* and *B* are the linguistic labels defined by fuzzy sets on the range of all possible values of *x* and *y*, respectively. The IF part of the rule "*x is A*" is called antecedent and the THEN part of the rule "*y is B*" is called consequent. The antecedent may integrate several inputs using logical AND and OR operators. For estimating the probability of a Probable-LM cluster based on the *NDLMI*, *RMSEyz*, and *Linearity*, one of the IF–THEN fuzzy rules might be the following: *IF NDLMI is High AND Linearity is High AND RMSEyz is Low THEN LM-Probability is true-LM*. Moreover, the

Table 1
Specification of the MLS systems in this research.

		I-85 BL	EWPkwy
LiDAR	Brand	Optech	Riegl
	Model	SG1	VMX450
	Single / Dual laser	Dual	Dual
	Measurement rate	600 kHz/ sensor	1.1 MHz
DMI	Brand	Applanix	Applanix
	Model	HS35F	BEI HH5
IMU	Brand	Applanix	Applanix
	Model	FMU P301	AP50
	Roll/pitch accuracy	0.005°	0.005°
	Heading Accuracy	0.015°	0.015°
Camera	Type	Point Grey 360°	NIKON/RIEGL
	No. of Cameras	6	4
	Focal Points of Cameras	N/A	2 front, 2 rear
	Resolution	5 MP	5 MP
Vehicle Mounted GPS/ GNSS	Brand	Trimble	Trimble
	Model	AT1675-540TS	Zepher model 2
	Accuracy	0.02′ H; 0.04′ V	10 mm

Table 2
Statistical information and parameter setting in preprocessing step.

	Dataset	I-85 BL	EWPkwy
Parameters	Overlap(m)	1	1
	W_r(m)	10	5
	W_l(m)	20	10
Total number of points		508,336,599 (100%)	428,321,603 (100%)
Number of deleted points after buffering		78,863,763 (16%)	153,246,160 (36%)
Number of extracted points in all sections		429,472,836 (84%)	275,075,443 (64%)
Number of sections		842	633
Number of valid sections		792	631
Section length (m)	Min	9.89	9.50
	Max	24.52	29.50
	Mean	19.77	14.96

strength of these IF-THEN rules can be considered by a weight value in the range of [0, 1]. For instance, since the above example is very reliable, this rule can get a weight equal to one. Here, 42 rules are constructed in order to make a decision about a LM cluster based on the input variables. For example, four selected sample rules and their considered weight are as follows:

- If (*NDLMI* is *Extreme*) and (*LMI* is *Extreme*) then (*LM-Probability* is *true-LM*) (1)
- If (*NDLMI* is *High*) and ($RMSE_{yz}$ is *Low*) and (*Linearity* is *Medium*) and (*LMI* is *High*) then (*LM-Probability* is *true-LM*) (1)
- If (*NDLMI* is *Medium*) and ($RMSE_{yz}$ is *Low*) and (*Linearity* is *Low*) and (*LMI* is *Medium*) then (*LM-Probability* is *false-LM*) (0.7)
- If (*NDLMI* is *Low*) and (*LMI* is *Low*) then (*LM-Probability* is *false-LM*) (1)

The result of all triggered rules can be aggregated using a sum function. Therefore, a defuzzification process has to be implemented in order to obtain a deterministic value. The most conventional defuzzification method is to calculate the center of gravity, which determines the center of the area under the aggregated output function (Zimmerman, 1996). Considering n as the number of elements of the sampled membership function, and $\mu(s)$ as the membership grade of the measurement, the center-of-gravity (CoG) method for discrete data can be calculated from Eq. (8).

$$CoG = \frac{\sum_{s=1}^{n} s.\ \mu(s)}{\sum_{s=1}^{n} \mu(s)} \tag{8}$$

The obtained center-of-gravity gets a value in the range of [0, 1]. The probable-LM cluster will be considered as the true-LM cluster if this value is greater than 0.5. Conversely, in the case of small probability, the cluster will be labeled as false-LM points and will be eliminated. In this paper, the theoretical background of the Fuzzy Inference System is not fully described. Readers are referred to Cox et al. (1998), Klir and Yuan (1996) and Yuan and Shaw (1995) for further discussion on fuzzy systems.

All extracted probable-LM clusters in each section are analyzed in the fuzzy LM detection system based on their input descriptors and false-LM clusters are deleted from the section. The output of this step is the extracted true-LM clusters in the candidate section that is refined in the next step.

3.5. Regularization

Two regularization processes are performed on each true-LM cluster: gap-filling and realignment. Since the LM points and the non-LM points have been differentiated based on the intensity information, a number of LM points with low intensity might not have been located in the probable-LM cluster. This may cause gaps in the LM. In the gap-filling process, all points inside the boundary of the current LM points are added to the LM points. Here, the boundary is detected based on the α-shape concept, developed by Edelsbrunner and Mücke (1994). For a defined set of points S, the α-shape for $\alpha = 0$ is identical to the original set of points S, and, $\alpha = \infty$, results in all triangles in an α-shape, which is equal to the convex hull of S (Cholewo and Love, 1999). A proper value of α can be considered based on the minimum distance between lane pavement markings on the road and the resolution of the MLS point cloud. Eliminated points of "non-LM" clusters located inside the

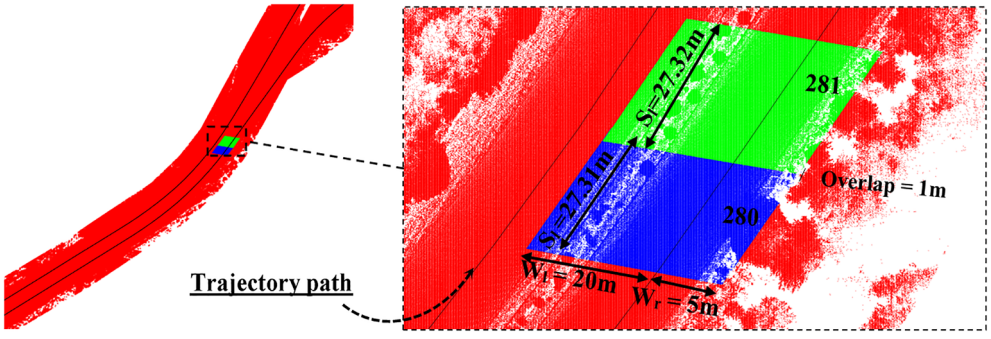

Fig. 6. The result of preprocessing step for two sample section. Red points are the omitted points after buffering. Blue and Green points are extracted points for two sample sections of 280 and 281 from I-85 BL dataset.

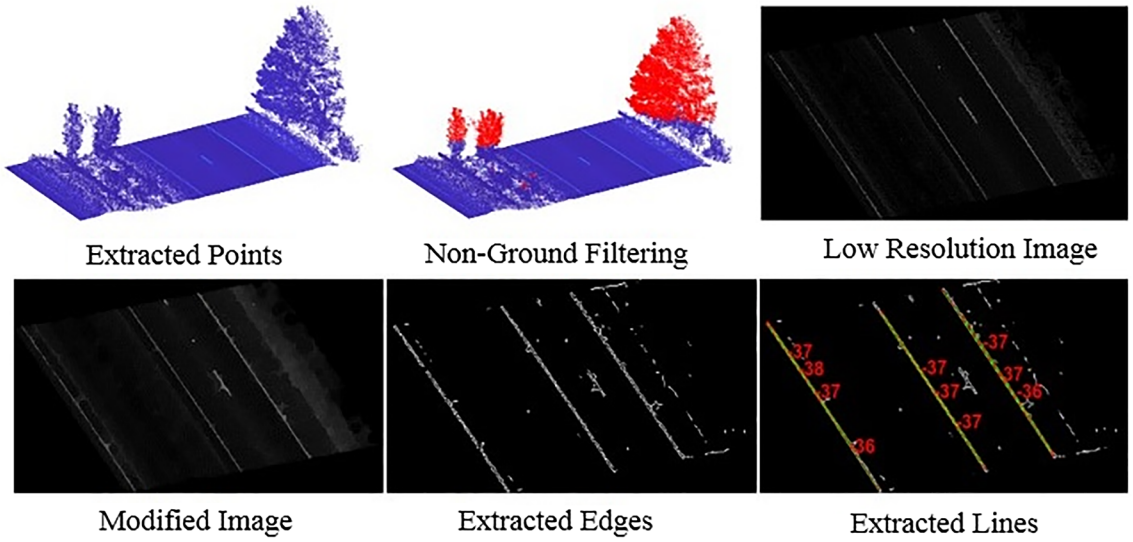

Fig. 7. Initial non-ground filtering and section alignment step for the sample 280 from I-85 BL.

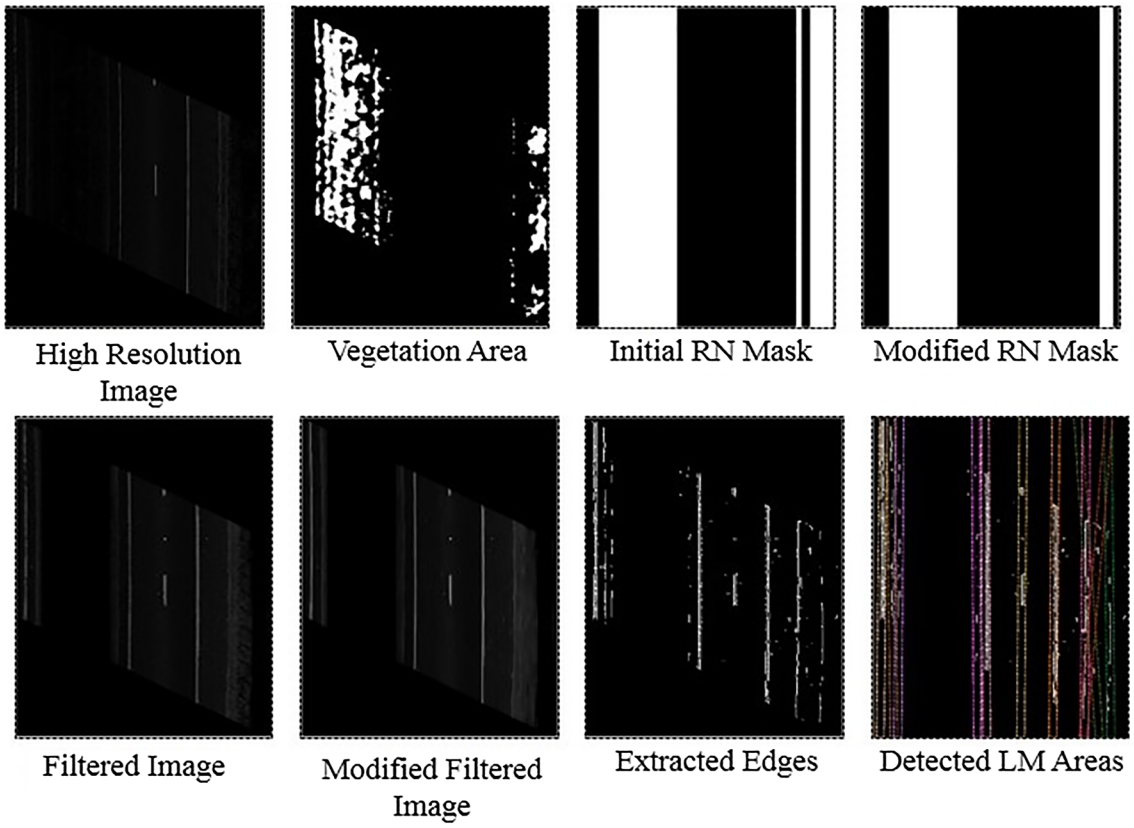

Fig. 8. Lane markings area detection process for the sample section 280.

obtained boundary based on this α value are added back to the LM cluster. After filling the gaps in each LM, the true-LM points can be transferred back to their original coordinates in the realignment step. In this case, a rotation mapping on each section must be performed based on its main direction angle. To avoid this computation cost, keeping the original coordinates of the points may be helpful. Moreover, the repetitive points in the overlapping areas are eliminated from the final lane marking points.

4. Experiments and results

4.1. Dataset

The efficiency of the proposed algorithm and methodology was evaluated with two point clouds, which were collected by several MLSs. The first point cloud includes 508 million points, which were collected along 15.6 km of Interstate 85 Business Loop (I-85 BL) in Spartanburg, South Carolina, USA. The second point cloud includes 428 million points collected along 9.5 km of East West Parkway (EWPkwy) in Anderson, South Carolina, USA. An overview of each study area is

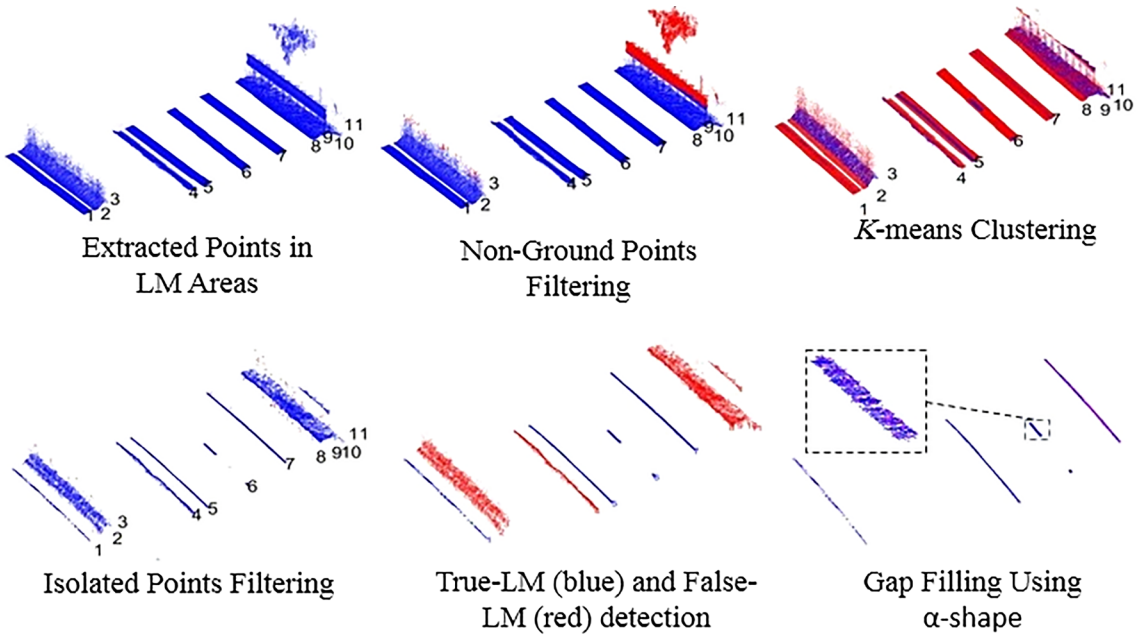

Fig. 9. LM areas detection, LM detection and enhancement steps for the sample section 280 from the I-85 BL dataset.

Table 3
The result of true-LM detection step using the fuzzy LM detection system for the sample 280 from I-85 BL dataset.

LM Number[a]	1	2	3	4	5	6	7	8	9	10	11
NDLI	0.73	0.43	0.47	0.41	0.72	0.77	0.72	0.42	0.39	0.35	0.37
RMSEyz	0.12	0.71	0.99	0.41	0.10	0.10	0.09	0.49	2.82	1.34	0.91
Linearity	109.39	64.37	43.40	63.19	133.21	90.39	132.16	59.13	24.51	30.56	26.68
LMI	**654.98**	38.96	20.50	64.20	**920.21**	**672.89**	**1000.60**	51.13	3.37	8.05	10.87
Fuzzy Output	**0.68**	0.34	0.33	0.33	**0.67**	**0.71**	**0.68**	0.33	0.31	0.33	0.32

[a] The LM numbers are shown in Fig. 9.

Table 4
Statistical information of the gap-filling step considering $\alpha = 0.1$.

	I-85 BL	EWPkwy
Average number of LM points in each section	34,582	147,250
Average number of added points by α-shape	508	1489
Percentage of the added points	2.22%	1.01%

shown in Fig. 5. Both datasets include a variety of geometric design elements, such as merging and diverging sections, tangent sections, horizontal and vertical curves, and bridges. Data collection specifications of the MLS vendors are summarized in Table 1.

4.2. Data processing

In this subsection the processes and parameter settings of different steps of the algorithm are presented and discussed. In the sectioning stage, the interval between two consecutive images were considered as section length (S_l) with 1 m overlap (S_o) between tiles. Each observed section spanning less than 2 m was considered "short" and was merged with its preceding section. A final sectioning and buffering of two sample sections of I-85 BL are shown in Fig. 6.

The parameter settings and basic statistics of the datasets are illustrated in Table 2. In the case of extracting two sample sections with IDs 280 and 281 from the I-85 BL dataset unnecessary points are shown in red in Fig. 6 and have been eliminated from the point cloud. After sectioning the whole data about 16% and 36% of points were deleted in I-85 BL and EWPkway, respectively.

In the initial non-ground filtering, points were filtered to include those located 2 m above or below a fitted plane to the road surface. The filtered points for section 280 are depicted by the red color in Fig. 7. An intensity image of the candidate sections was generated and modified using *Closing* morphological operators by applying a disk-shaped SE with a 2-pixel radius using a 0.1 m pixel size during section alignment. Edge images were extracted with the *Sobel* edge detection filter, and all lines were extracted through the HT algorithm. The search space of the HT algorithm was limited based on a calculated azimuth of the section Az_t as $[Az_t\text{-}10^\circ, Az_t+10^\circ]$, where Az_t can be obtained using trajectory information. The minimum number of pixels required to accept an extracted line was considered equal to the approximate length of dashed lane marks, about 2 m or 20 pixels on the image. For example, in the sample section 280, the HT algorithm extracted all of the lines on the edge with length greater than 20 pixels and a difference between their azimuth and the section azimuth of less than 10°. Among the extracted lines, the most frequent azimuth was assigned as the primary direction of the section which was -37° for sample section 280. The step-by-step process of section alignment of this sample section is shown in Fig. 7.

All points were mapped in the Y direction according to the rotation matrix, and a new intensity image with 0.05×0.05 m pixel size was generated. The generated grid format for the selected sample section is shown in Fig. 8. The vegetation areas were filtered using an RN-mask obtained from the RN information. For this purpose, an RN image was generated to create the RN-mask and modified using *Dilation* and *Erosion* morphological operators by considering $v = 113$ (equal to the half of the length of this section) and $h = 10$ in Eqs. (1) and (2). After filtering the intensity image with this mask, the resulting image was modified using the *Closing* operator and based on a disk-shaped SE with

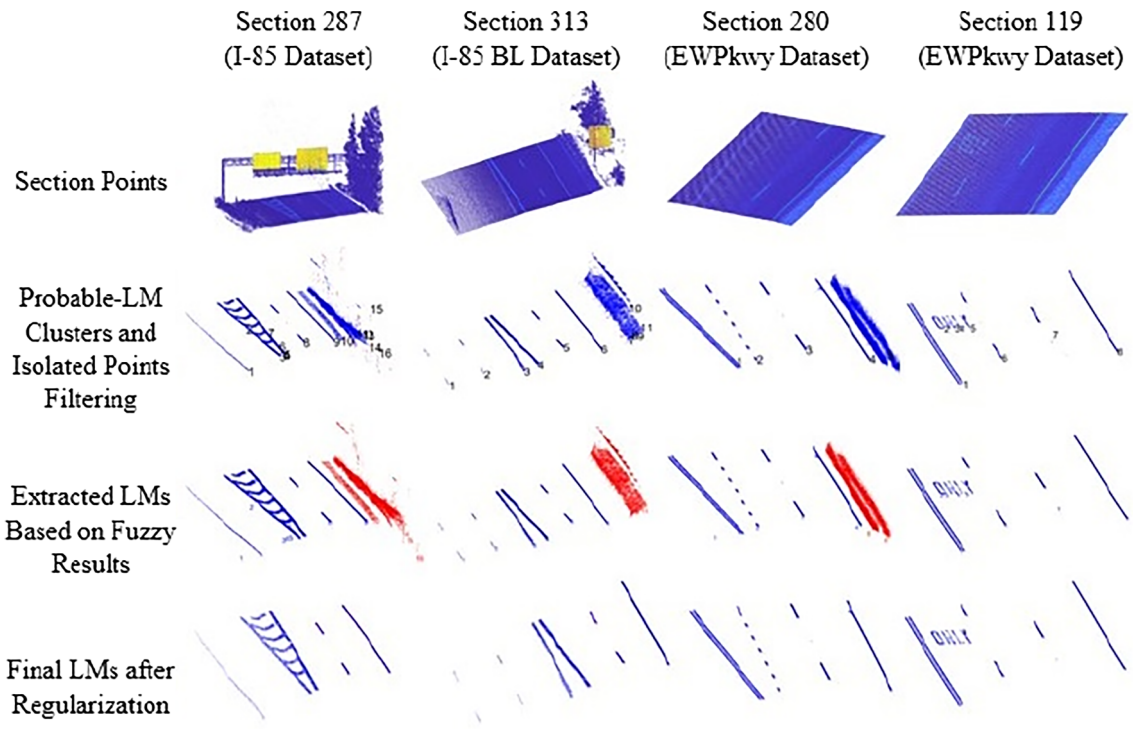

Fig. 10. The results of LM detection process using the proposed method for three selected sample sections.

Table 5
Extracted values, and the result of fuzzy LM detection system for sample sections shown in Fig. 10.

Section ID	LM Number	1	2	3	4	5	6	7	8	11	12	13	14	15	16
287 (I-85 BL)	NDLMI	0.7	0.8	0.7	0.8	0.8	0.7	0.8	0.8	0.4	0.4	0.4	0.4	0.6	0.5
	RMSEyz	0.3	0.2	0.2	0.2	0.2	0.2	0.1	0.1	1.2	3.1	3.5	2.8	3.2	1.7
	Linearity	101.7	30.3	43.0	39.3	44.8	47.4	38.0	137.2	41.9	33.4	40.6	39.2	29.0	20.6
	LMI	244.9	116.9	158.6	153.5	154.1	195.7	202.8	960.7	14.4	4.4	4.6	6.0	5.2	5.7
	Fuzzy Output	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**	0.3	0.3	0.3	0.3	0.4	0.3
313 (I-85 BL)	NDLMI	0.8	0.8	0.7	0.7	0.8	0.7	0.4	0.3	0.4					
	RMSEyz	0.1	0.1	0.1	0.1	0.1	0.0	0.2	0.2	0.2					
	Linearity	160.4	165.3	88.5	77.1	128.1	144.3	51.1	83.5	49.1					
	LMI	1805.6	2642.5	783.3	499.3	2034.2	2361.1	95.2	**122.8**	93.3					
	Fuzzy Output	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**	0.3	0.3	0.3					
280 (EWPkwy)	NDLMI	0.7	0.8	0.7	0.7	0.4	0.1								
	RMSEyz	0.1	0.0	0.0	0.0	0.2	0.2								
	Linearity	40.3	170.0	189.2	133.9	37.7	40.1								
	LMI	307.7	3581.6	3900.3	3070.5	96.3	24.0								
	Fuzzy Output	**0.7**	**0.7**	**0.7**	**0.7**	0.3	0.3								
119 (EWPkwy)	NDLMI	0.7	0.8	0.8	0.8	0.8	0.8	0.7	0.7						
	RMSEyz	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.0						
	Linearity	41.6	10.4	3.8	5.8	8.2	176.2	33.3	158.4						
	LMI	465.4	198.9	37.5	76.3	131.0	3760.5	689.8	2881.9						
	Fuzzy Output	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**	**0.7**						

a 2 pixels radius. Next, the vertical *Sobel* edge extraction operator was applied to extract the edge pixels. All candidate LM areas can be extracted using the HT algorithm considering $[-15^{\circ}, +15^{\circ}]$ with a 0.1° sampling interval as the angle search interval. In sample section 280, 11 LM areas were detected that are shown in Fig. 8.

A 3D view of all extracted points from LM areas and the filtered non-ground points for the sample section 280 are shown in Fig. 9. The figure shows that MLS objects such as tree branches and guardrail not of interest were successfully eliminated after the original non-ground filtering process. The *k*-means clustering algorithm was used to consider 5 initial values for the two cluster centers (probable-LM and non-LM clusters) based on Eq. (3). Higher intensity clusters are labeled as probable-LM cluster that are depicted in blue in Fig. 9 and the detected isolated points in each probable-LM cluster are shown in red. In this case, dX, dY, dZ and the minimum number of points were considered 0.1 m, 2 m, 0.1 m and 5, respectively. The radiometric and geometric features and the output of the fuzzy LM detection system for the probable-LM clusters are shown in Table 3. In this table, there is a spectacular similarity between the descriptors in the 1st, 5th, 6th and 7th probable-LM clusters, which are true-LM clusters. Consequently, the fuzzy output of these clusters is far from the other clusters. A comparison between the values shown in Table 3 and the fuzzy output of the probable-LM clusters for the sample section is an indication that the process works well in terms of detecting true-LM clusters. These clusters were further regularized in the gap filling process through the α-shape algorithm assuming $\alpha = 0.1$. The newly added points in the gap filling process are highlighted in pink in Fig. 9.

The potential gaps within the true-LMs are filled through regularization with a 0.1 α value. This resulted in 2.22% of the extracted LM points in I-85 BL and 1.01% in EWPkwy datasets being added, as

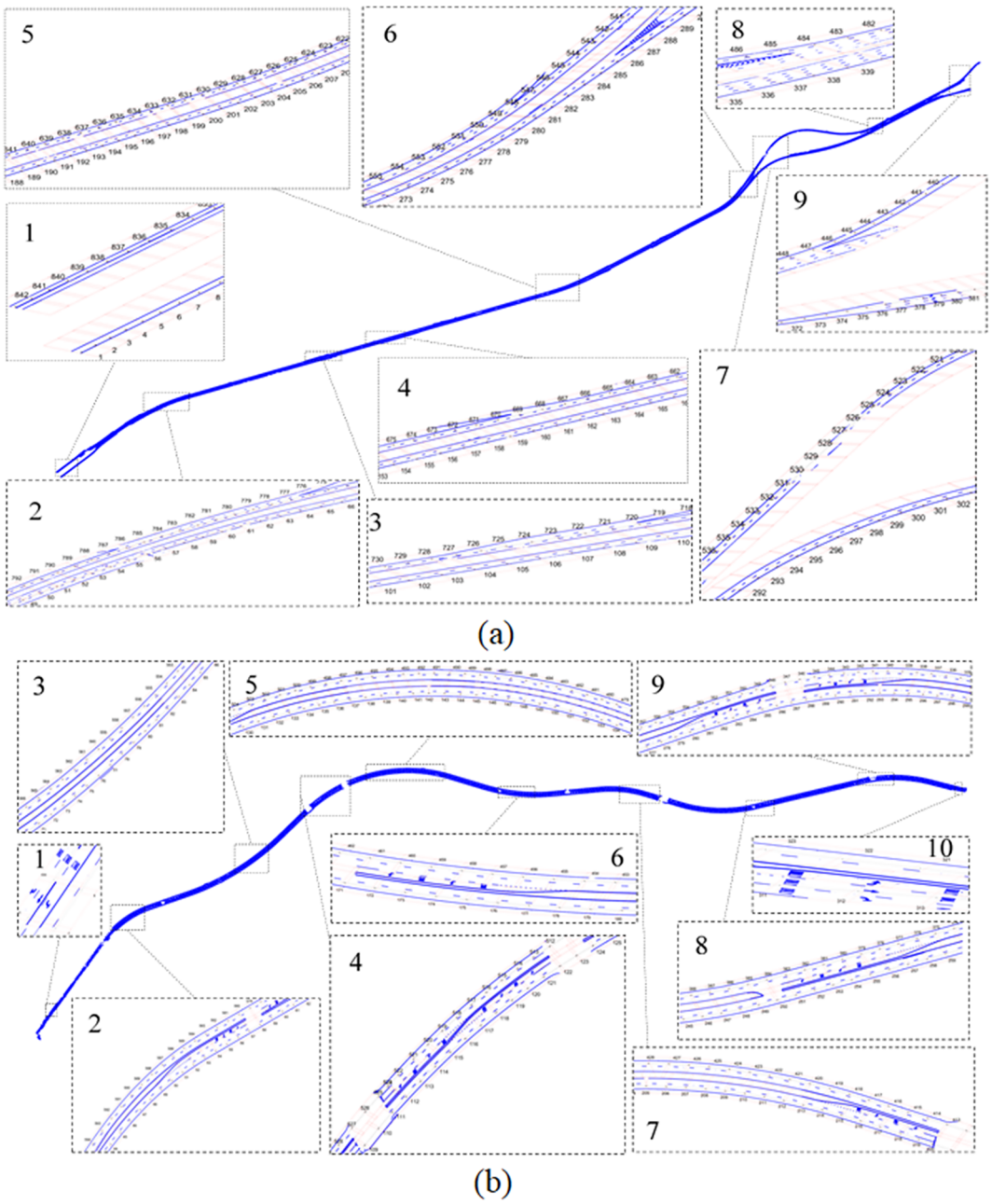

Fig. 11. Final extracted LM from MLS point cloud using the proposed fuzzy LM detection system. (a) I-85 BL dataset; (b) EWPkwy dataset.

illustrated in Table 4. The repetitive points due to the overlap of the sections are taken into account in this table. This α was determined through a test of several α values which gave the best results, particularly in the zebra-shaped markings and word markings.

4.3. Results

The algorithm was performed on all sections for both datasets and the results of four sample sections are shown in Fig. 10. As can be seen from this figure, the algorithm has successfully extracted all lane marking points.

Table 5 illustrates the value of the geometric and radiometric features and the fuzzy result of these sections. As shown in this table, the LMI value of the 8th probable-LM cluster of section 313 from the I-85 BL dataset is 122.8 even though it is not a true-LM cluster. Conversely, in the 3rd and 4th probable-LM clusters of section 119 from the EWPkwy dataset the LMI value is small even though it is a true-LM cluster. This uncertainty between true-LM clusters and false-LM clusters can also be observed for other descriptor values in these four sections. Each value was successfully handled through the application of all values in the fuzzy inference system.

The extracted LMs for all sections for both datasets are shown in Fig. 11 along with zoomed views of sections with varied markings including text markings, turn arrows, turn bay markings, dashed longitudinal markings, and painted islands.

4.4. Accuracy assessment

To evaluate the reliability of the proposed method, a comparison was made between the extracted LMs from the algorithm and the manually extracted LMs. Considering the LM process as a two-class classification problem, in which the outcomes are labeled as LM point or non-LM point, there are four possible outcomes. If a point from the method is labeled as LM and is included in a manually extracted LM,

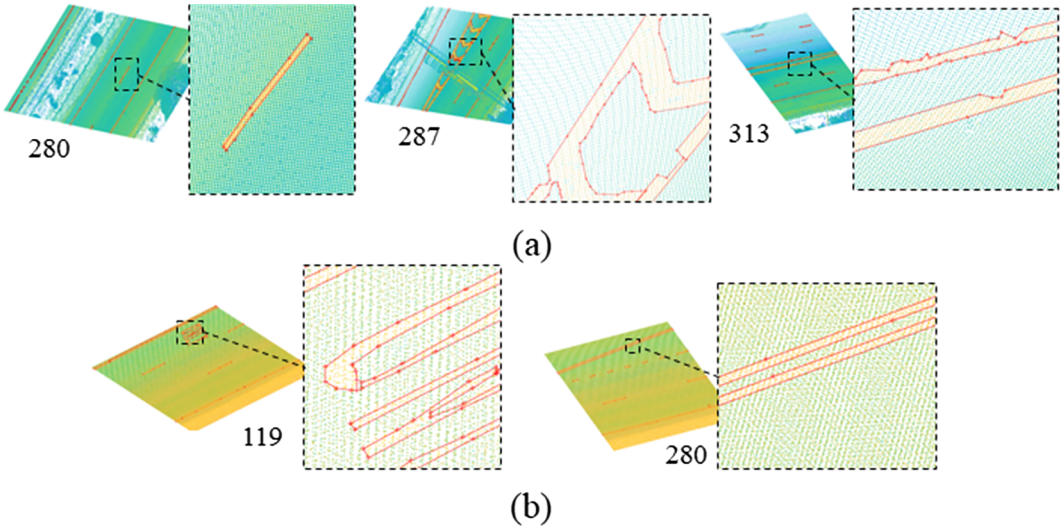

Fig. 12. Manually extracted LMs from point cloud; (a) sample sections of the I-85 BL dataset; (b) sample sections of the EWPkwy dataset.

Table 6
Evaluation of the extracted LM points in the five selected sample sections shown in Fig. 12.

	S.ID.	Length (m)	#LM	#Pts.	FP	FN	TP	TN	TPR	FPR	F1	ACC
I-85 BL	280	27.3	11	672,152	10,008	1344	1643	659,157	0.859	0.002	0.870	0.996
	287	28.47	16	698,211	16,336	789	3215	677,871	0.836	0.001	0.891	0.994
	313	24.21	11	529,793	12,746	1690	1657	513,700	0.885	0.003	0.884	0.994
EWPkwy	119	20.40	8	434,514	10,039	686	1674	422,115	0.857	0.002	0.895	0.995
	280	18.34	6	434,741	9139	1371	995	423,236	0.902	0.003	0.885	0.995
Average									**0.868**	**0.002**	**0.885**	**0.995**

Table 7
Performance of the fuzzy inference system in distinguishing true-LM clusters from false-LM clusters (25 km).

S.ID.	#P.L.	FP	FN	TP	TN	TPR	FPR	F1	Accuracy
I-85 BL	9984	3516	9	34	6425	0.990	0.001	0.994	0.996
EWPkwy	4360	2764	25	11	1560	0.996	0.016	0.994	0.992

then it is classified as a True Positive (TP), otherwise it is called a False Positive (FP). Conversely, for a non-LM point in a manually processed section, a True Negative (TN) and a False Negative (FN) will occur if the algorithm labels the point as non-LM point and LM point, respectively. Based on these parameters, a number of derivations such as True Positive Rate (TPR), False Negative Rate (FNR), Accuracy, and F1-score can be calculated (Powers, 2007; Zaboli et al., 2019). These parameters can also be calculated to assess the performance of the fuzzy LM detection step by visually checking the correctness of the detected true-LMs.

In this study, all LM points of the five sample sections (280, 287 and 313 from I-85 BL, and 119 and 280 from the EWPkwy) were manually extracted by plotting top-view of the section and considering a polygon around each LM area in the section. This process was labor intensive and time-consuming, whereas extracting pavement LMs take fifteen minutes for each section on average under the proposed new method. Fig. 12 shows the extracted polygons of the LM areas for sample sections overlaid on the top view of the point cloud.

Manually extracted LM points were compared to the algorithm LM points, and the evaluation measures of each section are shown in Table 6. The high value of the F1-score, TPR and ACC measures in all sections plus the low rate of the FPR value indicates that the proposed method to extract traveling LM works well. The maximum and minimum TPR measures among these five sample sections were reported 0.902 for section 280 from the EWPkwy dataset and 0.836 for section 287 from I-85 BL dataset, respectively. Section 119 from EWPkwy performed the best based on its F1-score of 0.895 while section 280 from the I-85 BL dataset had the lowest F1-score. The slight difference between TPR, FPR, F1-score and ACC in these five sections shows the stability and reliability of the performance of the proposed method. Although FPR and ACC parameters are important in the evaluation of any classification, they may not show the exact differentiation between the performance of the algorithm on different sections due to a large number non-LM points.

The performance of the fuzzy LM detection system was checked by observing the obtained true-LMs along with the route. All obtained true-LM clusters were plotted (Fig. 11) for accuracy and mistaken decisions. The true-LM clusters were counted to estimate the confusion matrix. Table 7 shows the obtained confusion matrix and the estimated parameters for both datasets (25 km). As can be observed in this table, the obtained parameters for both datasets indicate that the proposed fuzzy LM detection system can successfully detect true-LM markings clusters at a very high level of accuracy.

Considering the lengthy datasets tested in this study, the evaluation parameters, such as TPR and F1-Score shown in Tables 6 and 7, indicate the efficiency of the proposed algorithm in extracting LM. The reported F1-Score and TPR in Table 7 of nearly 99% for fuzzy LM detection shows a great deal of promise. However, points with low intensity values may not be detected as probable-LM clusters due to the weight of *k*-means clustering. The manually extracted LM were obtained by considering a number of polygons around the LM and selecting all of the points inside those polygons (see Fig. 12). This must be considered when evaluating the performance of the system.

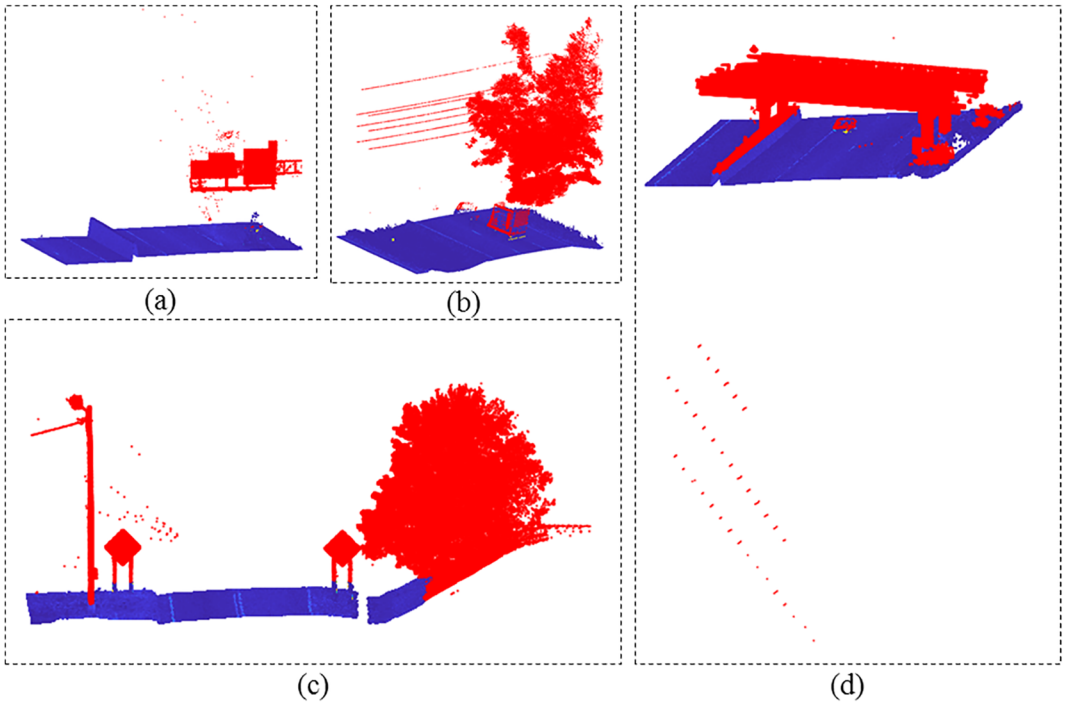

Fig. 13. The performance of the proposed initial object filtering method in four sample sections presented many challenges including: (a) Information sign above the road; (b) tree branches, wires, and cars; (c) power pole, traffic signs, tree branches, noisy points below and above ground surface, and high-level ground surface on the road side; (d) non-planer intersection, noisy points below and above ground surface, tree branches, and cars.

5. Discussion

5.1. Input data

This method assumes that MLS vehicle trajectory data is available. If this is not the case, a roadway centerline downloaded from a digital map or a manually drawn trajectory can be substituted. Further, the proposed method uses the RN value for vegetation area filtering in order to increase the computation speed during the LM areas detection process. This information may not be an available attribute of the MLS point cloud. The necessity of the RN values for the point clouds was tested in many sections. It was observed that in the sections with vast vegetation area, a large area was filtered by the RN-mask and the number of extracted lines using HT algorithm was reduced. When the RN value was not used for those sections, a few extra LM areas were detected by the HT algorithm. However, all of the extracted probable-LMs in these areas were successfully detected as false-LM in the fuzzy inference system. Therefore, it can be concluded that the RN value is not essential in our proposed algorithm, and in the case where RN value is not available the algorithm still remains effective and provides promising results.

5.2. Selection of parameters

Object filtering in this research was performed in two steps: the coarse filtering in the preprocessing step and the fine filtering in the probable-LM detection phase. Since the proposed initial filtering works based on a few selected points in six areas (see Fig. 1), it is a fast and easy-to-implement algorithm. In this case, the predefined threshold (2 m above or below) was selected based on performing various tests on the sections. Note that the number of point selection areas can be increased depending on the road width. Fig. 13 shows the performance of the initial non-ground filtering based on this threshold. As can be seen from this figure, the proposed initial filtering methods have successfully filtered unwilling objects as well as noisy points.

The statistical information of the preprocessing step in Table 2 shows that a large number of unneeded points (16% in the I-85 BL dataset and 36% in the EWPkway dataset) were filtered in the buffering step. The buffering space parameters (W_r and W_l) were selected based on the previous knowledge about these datasets. For a different project, they should be set based on the maximum road width of the survey path to include all LMs. Limiting the parametric space of the HT algorithm helps to increase the processing speed of both section alignment and initial LM areas detection steps. In the section alignment step, the azimuth of the trajectory line could influence the results. Therefore, if the available trajectory line is not smooth nor parallel to the road's LMs, a larger threshold must be considered. Otherwise, extracting the lines in all possible directions, ± 90°, may need further verification, especially in sections with damaged LMs. In this research, ± 10° was an appropriate value due to the smoothness of the trajectory lines. If trajectory smoothness is a concern, the use of a ± 15° angle search interval may help to compensate for possible trajectory alignment errors and can facilitate the extraction of LMs at road exits.

The number of clusters in the *k*-means clustering algorithm may not be a challenging parameter because the problem of extracting probable-LM clusters has been considered as a thresholding problem thus, 2 clusters is a perfect choice. Moreover, as mentioned previously, the α value in the regularization step was determined through a test of several α values which gave the best results. Therefore, due to the standard and fixed width of the LMs, it is expected that this value provides promising results for other projects. Note, this parameter fills the gaps between points of a specific LM. Therefore, in the case of studying and assessing the degree of damage of LMs, this step may not be necessary because the damaged area inside the lane markings may be ignored.

5.3. Descriptors analysis

A strong correlation between the intensity values from mobile laser scanning and pavement marking retro-reflectivity was found by Olsen et al. (2018). Since the retro-reflectivity of lane markings diminish over time detecting lane markings based only on intensity may not be possible. However, by analyzing geometric and radiometric descriptors in

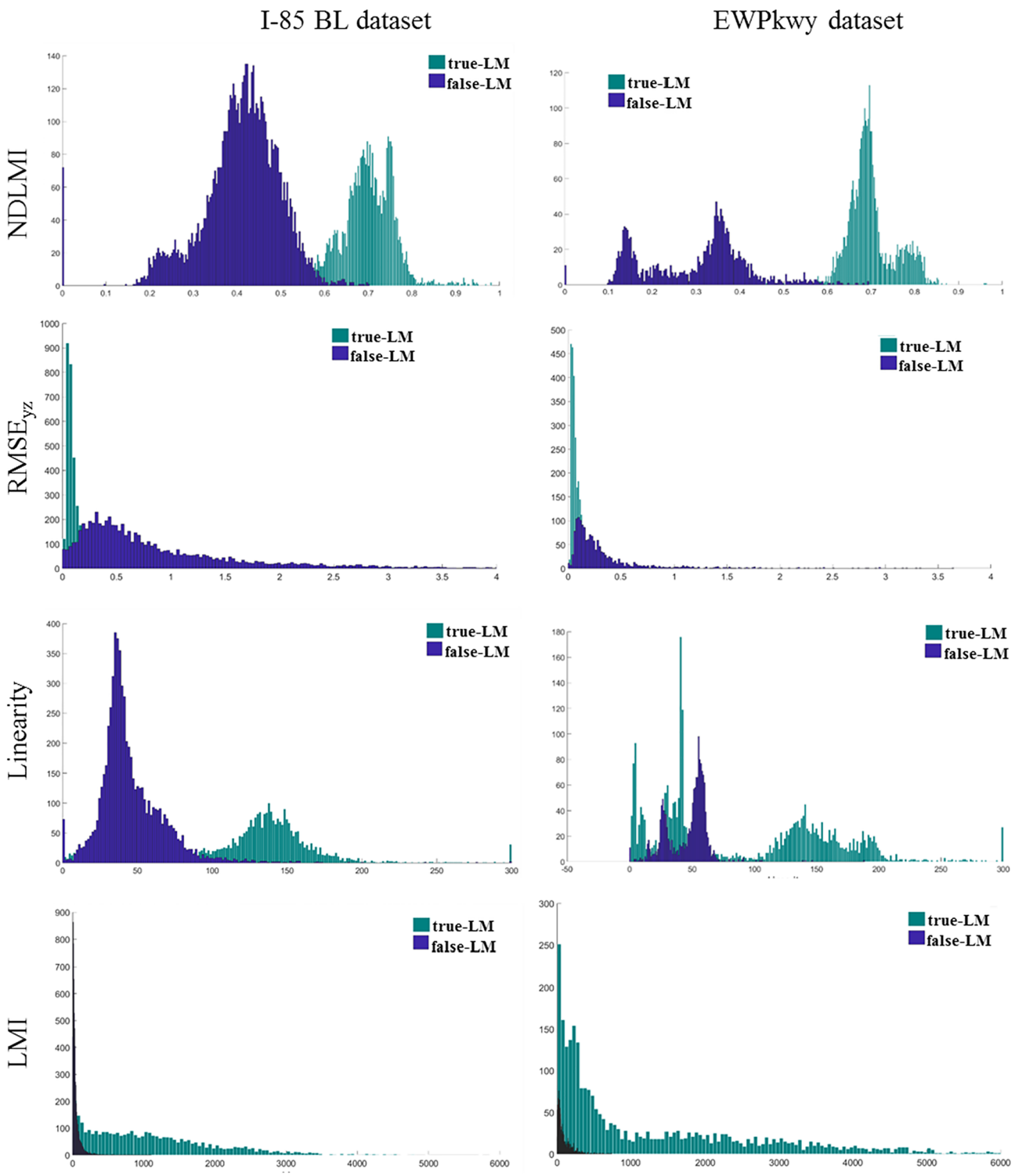

Fig. 14. Distribution of applied descriptors for true-LM and false-LM clusters.

the proposed fuzzy inference system low quality lane markings were also extracted successfully. These descriptors as input vectors are important to the performance of the applied fuzzy LM detection system. Fig. 14, depicts the distribution of these descriptors for the final detected true-LM and false-LM clusters.

As shown in Fig. 14, there is a clear difference between *NDLMI*, in both datasets, and the *Linearity*, in the I-85 BL dataset. However, the *Linearity* of the true-LM points in the EWPkwy dataset can be seen along most of the range of this descriptor. This may be due to the large number of short probable-LM clusters in this dataset where their *Linearity* value is small (see the *Linearity* values of section 119 in the EWPkway dataset in Table 5). The trend of the $RMSE_{yz}$ and the *LMI* values in the I-85 BL dataset are approximately the same as that of the EWPkwy dataset.

Table 8
Quantitative evaluation results of different lane marking extraction methods.

Methods	Datasets	Number of points (million points)	TPR (%)	F1-Score (%)
Kumar et al. (2014)	140 m	–	88	–
Guan et al. (2014)	Data#1: 105 m Data#2: 63 m	Data#1: 8.4 Data#2: 5.4	83	89
Soilán et al. (2017)	1510 m	69	82–90	92–93
Yang et al. (2017)	Data#1: 5.3 km Data#2: 79.8 km	Data#1: 377.3 Data#2: 564.2	94.5	> 90
Soilán et al. (2018)	Data #1: 2.5 km Data #2: - Data 3#: -	Data #1: 129 Data #2: 82 Data #3: 50	91.7	93.9
Wen et al. (2019)cGAN-based	Data #1: - Data #2: -	Data #1: - Data #2: -	76–82	79–86
Wen et al. (2019)U-net-based	Data #1: - Data #2: -	Data #1: - Data #2: -	81–87	85–92
Jung et al. (2019)	Data #1: 460 m Data #2: 2.1 km Data #3: 4.7 km	Data #1: 19.6 Data #2: 115.0 Data #3: 96.0	74–95	89–97
Our method	Data #1: 15.6 km Data #2: 9.5 km	Data #1: 508 Data #2: 428	84–90	87–90

Table 9
Processing time of executing the proposed method on each dataset.

	I-85 BL	EWPkwy
Point Cloud Volume (GB)	13.2	11.1
Sectioning time (h)	2.68	1.34
Average LM extraction time per section (s)	9.4	6.05
Total LM extraction time (h)	2.1	1.06

5.4. Comparative study

Table 8 illustrates the size of the dataset and quantitative evaluation results in comparison to other studies. According to the higher TPR, the proposed method is superior to most of the previous similar studies. Note, in our study as well as Guan et al. (2014), the informative road markings (text, arrows, zebras, etc.) have also been extracted from the test area, while not included in other studies. In our method, the LM areas are detected using HT algorithm which extracts lines from the intensity image. Therefore, every linear LM along the road can be detected by our proposed algorithm. For informative markings such as text and elongated turning arrows, the edge extraction algorithms can extract their corresponding edges, and consequently, the HT algorithm extracts the LM areas around them. Since a buffer is considered around each line, the entire text, or turn arrow is extracted. Moreover, in a number of studies such as Wen et al. (2019) the extracted road markings were classified into different classes based on their types. Although, this information may be useful for many applications, the main goal of this research was to extract the road markings.

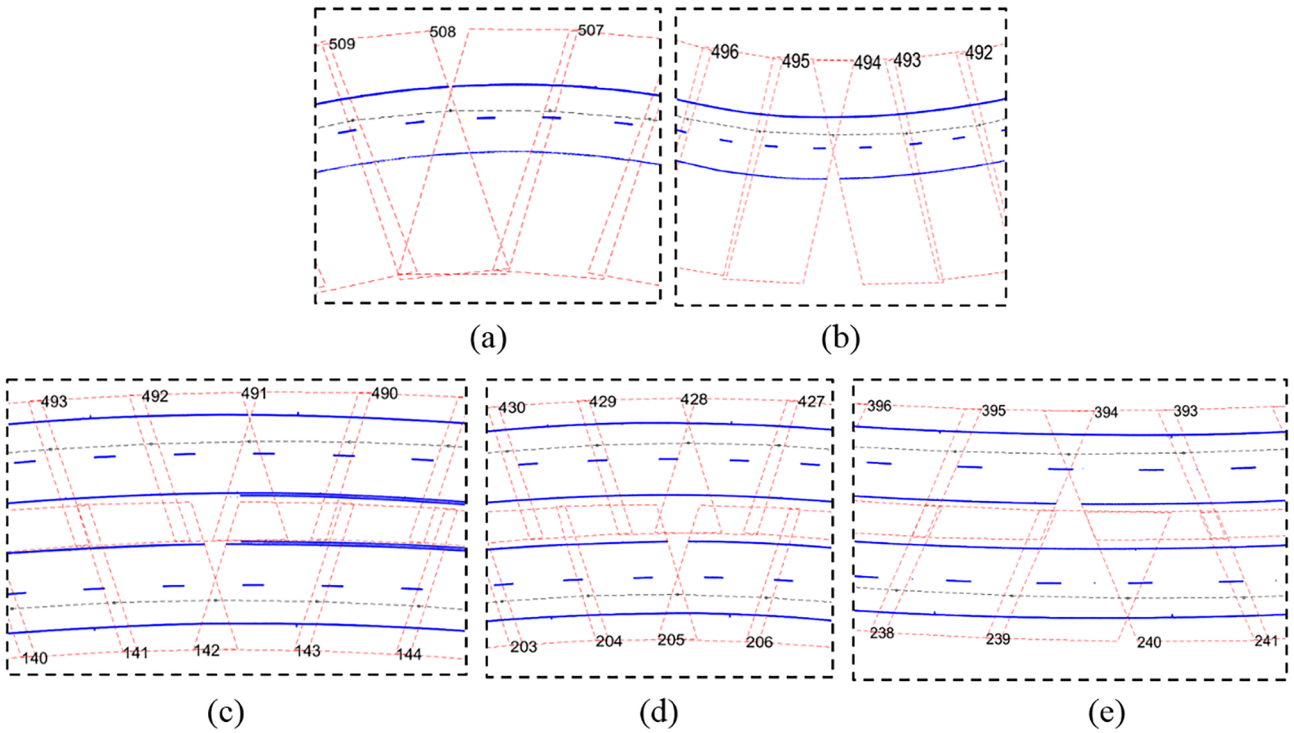

Fig. 15. Appeared gap areas due to the tangent changing from a quadrant to another during sectioning. (a) and (b) Gap areas in I-85 BL data; (c), (d) and (e) Gap areas in EWPkwy data.

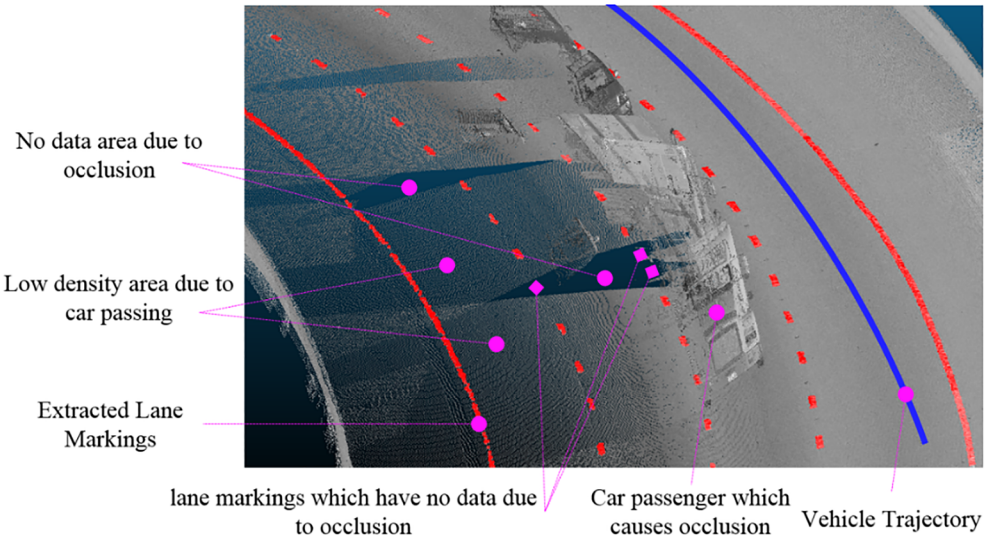

Fig. 16. Examples of occluded areas by passenger car.

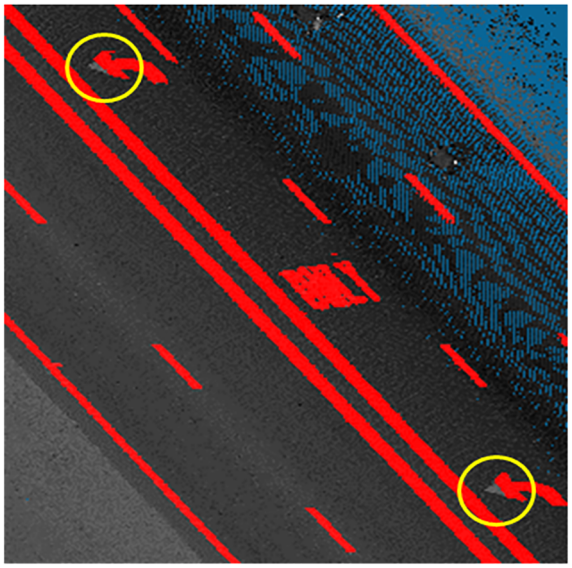

Fig. 17. Two samples of the short LM which is not extracted during HT algorithm.

Due to the variation of shape and retro-reflectivity of LMs along a road, an algorithm may not be efficient and robust throughout the entire path. Therefore, a lengthy corridor with various LMs in terms of shape and reflectivity would be vital for testing purposes. According to Table 8, the length and the size of the point cloud of our study are comparable to Yang et al. (2017) but not comparable to the previous LM extraction studies that used much smaller point clouds and shorter roadway lengths.

5.5. Processing time

Processing time is a key element in each algorithm. Our algorithm includes two main processing phases of sectioning and LM extraction. Table 9 illustrates the processing time of each phase for the two datasets. The algorithm was developed in MATLAB 2015a and executed on a normal computer with 8 GB of RAM and an Intel(R) Core(TM) i7-4770K CPU @3.50 GHz. Therefore, improving the hardware as well as the programming environment will decrease computation time.

One of the most important advantages of this algorithm is that the LM extraction process is largely unsupervised with only minimal user interaction. The majority of the parameters are constant for different projects, and parameter tuning may only be necessary for sectioning parameters (W_r, W_l, and S_o) due to differences in road width. With the addition of a user-friendly user interface and a simple to follow workflow from a human interaction standpoint, this method can be more efficient for LM extraction projects.

It should be noted that our algorithm detects and extracts all LM points based on their intensity. In this case, the low density points or the small width of the extracted LMs could be symptoms of their damage that can be considered for future studies. Nonetheless, manually extraction of such information based on visual inspection of the data may be much more time-consuming than the merely LM extraction.

For projects where trajectory data is not available, the required time for manually collecting a simulated trajectory or extracting a roadway centerline from a digital map should be added to the total processing time for this method. The user may add a post-processing step for further modification after executing the algorithm to achieve more reliable results which may require additional time.

Additionally, the potential for simultaneously processing various sections apart from each other is an added benefit. Thanks to the parallel processing techniques, processing speed can be significantly increased. Although the algorithm in this research was executed non-parallel, 1 to 2 h automatic processing is still unattainable with even short manual extraction time periods.

5.6. Challenges and future work

Even though an overlap distance was used between tiles on curve sections to account for gaps, due to the tangent changing from a quadrant to another during sectioning some short gaps were still discovered. This occurred in sections 494 and 508 in the I-85 BL dataset and caused a 4.3 m loss. This also occurred in sections 142, 205, 394, 428, and 491 from the EWPkwy dataset totaling a 7.6 m loss. Fig. 15 illustrates appeared gap areas in both I-85 BL and EWPkwy datasets. While, a fixed value for overlap distance was considered during the sectioning, an adaptive distance based on roadway curvature is most suitable to eliminate gaps.

Adjacent vehicles may either hide LMs or reduce the density of the points in a tile. Our proposed algorithm is unable to detect hidden lane markings where few or no points exist however it successfully detected lane markings in areas with low point density. The problem of occlusion can be minimized by controlling traffic during the data collection process. This can be accomplished by having trailing vehicles driving side by side to prevent other vehicles from passing the MLS vehicle.

Fig. 16 show sample occluded areas caused by a passenger car.

The proposed method can successfully extract information markings such as turn arrows. However, due to the minimum line length of the HT algorithm in the LM area detection step, very short lines, such as the tip of the direction arrows on the road surface, may not be extracted. Fig. 17 shows the detected LM points in sections 458 to 460 from the EWPkwy dataset overlaid on the main point cloud to illustrate this error.

6. Conclusions

This study proposed a state-of-the-art method to automatically extract highway pavement markings from MLS point clouds. In the proposed method, the trajectory data is applied in order to divide the whole dataset into a number of sections (e.g., tiles). In each tile, the candidate LM point groups identified as probable-LM are initially extracted. Additional processing using a fuzzy inference system gives a final extraction of true-LM points. The algorithm was tested on two datasets with lengths of 15.6 km and 9.5 km, and promising results were obtained. Reported average 88% F1-score and 87% TPR in five different sample sections from two study areas in addition to the impressive performance of the fuzzy LM detection system over the 25 km prove the reliability of the proposed algorithm in extracting road markings.

Using trajectory data and considering a buffer size in the sectioning step eliminated a large number of points. This step resulted in a dramatic reduction of the data volume and an increase of speed of the computation process. Switching between a point cloud format to raster format prevented the mistaken deletion of the LM points. Extracting the candidate LM areas, defining the *LMI* and *NDLMI* measures, and applying a fuzzy inference system to handle the uncertainty in recognizing true-LM clusters are key steps to the proposed method. Although the results of this algorithm were promising, the membership functions of the designed fuzzy LM detection system were assigned by an expert by observing many probable-LM clusters, which can be one of the drawbacks of the proposed method. Therefore, automatic detection of these membership functions by powerful algorithms such as Adaptive Neuro-Fuzzy Inference System (ANFIS) (Jang, 1993) may help to overcome this problem.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ai, C., Tsai, Y.J., 2016. An automated sign retroreflectivity condition evaluation methodology using mobile LIDAR and computer vision. Transp. Res. Part C: Emerg. Technol. 63, 96–113.

Bartels, M., Wei, H., 2010. Threshold-free object and ground point separation in LIDAR data. Pattern Recogn. Lett. 31, 1089–1099.

Chen, X., Kohlmeyer, B., Stroila, M., Alwar, N., Wang, R., Bach, J., 2009. Next generation map making: geo-referenced ground-level LIDAR point clouds for automatic retro-reflective road feature extraction. In: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, pp. 488–491.

Cholewo, T.J., Love, S., 1999. Gamut boundary determination using alpha-shapes, color and imaging conference. Society Imaging Sci. Technol. 200–204.

Cox, E., O'Hagan, M., Taber, R., O'Hagen, M., 1998. The Fuzzy Systems Handkbook with Cdrom. Academic Press Inc.

Edelsbrunner, H., Mücke, E.P., 1994. Three-dimensional alpha shapes. ACM Trans. Graphics (TOG) 13, 43–72.

Gatziolis, D., Andersen, H.-E., 2008. A Guide to LIDAR Data Acquisition and Processing for the Forests of the Pacific Northwest, Gen. Tech. Rep. Department of Agriculture, Forest Service, Pacific Northwest Research Station. PNW-GTR-768, Portland, OR: US.

Guan, H., Li, J., Yu, Y., Chapman, M., Wang, C., 2015. Automated road information extraction from mobile laser scanning data. IEEE Trans. Intelligent Transp. Syst. 16, 194–205.

Guan, H., Li, J., Yu, Y., Wang, C., Chapman, M., Yang, B., 2014. Using mobile laser scanning data for automated extraction of road markings. ISPRS J. Photogram. Remote Sensing 87, 93–107.

Habib, A., Ghanma, M., Morgan, M., Al-Ruzouq, R., 2005. Photogrammetric and Lidar data registration using linear features. Photogram. Eng. Remote Sensing 71, 699–707.

Hatger, C., Brenner, C., 2003. Extraction of road geometry parameters from laser scanning and existing databases. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 225–230.

Höfle, B., Mücke, W., Dutter, M., Rutzinger, M., Dorninger, P., 2009. Detection of building regions using airborne Lidar–a new combination of raster and point cloud based GIS methods. In: Proceedings of GI-Forum 2009—International Conference on Applied Geoinformatics, pp. 66–75.

Holgado-Barco, A., González-Aguilera, D., Arias-Sanchez, P., Martinez-Sanchez, J., 2015. Semiautomatic extraction of road horizontal alignment from a Mobile LiDAR system. Computer-Aided Civil Infrastruct. Eng. 30, 217–228.

Holgado-Barco, A., Riveiro, B., González-Aguilera, D., Arias, P., 2017. Automatic inventory of road cross-sections from mobile laser scanning system. Computer-Aided Civil Infrastruct. Eng. 32, 3–17.

Jaakkola, A., Hyyppä, J., Hyyppä, H., Kukko, A., 2008. Retrieval algorithms for road surface modelling using laser-based mobile mapping. Sensors 8, 5238–5249.

Jain, A.K., 2010. Data Clustering: 50 Years Beyond K-means. Pattern Recogn. Lett. 31, 651–666.

Jang, J.-S., 1993. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst., Man, Cybernet. 23, 665–685.

Jiang, H., 2017. Semi-automated Generation of Road Transition Lines Using Mobile Laser Scanning Data. University of Waterloo.

Jolliffe, I., 2011. Principal Component Analysis International Encyclopedia of Statistical Science. Springer, pp. 1094–1096.

Jung, J., Che, E., Olsen, M.J., Parrish, C., 2019. Efficient and Robust Lane Marking Extraction from Mobile Lidar Point Clouds. ISPRS J. Photogram. Remote Sensing 147, 1–18.

Klir, G.J., Yuan, B., 1996. Fuzzy Sets, Fuzzy Logic, And Fuzzy Systems: Selected Papers by Lotfi A. Zadeh. World Scientific Publishing Co., Inc.

Kumar, G., Patil, A., Patil, R., Park, S., Chai, Y., 2017. A LiDAR and IMU integrated indoor navigation system for UAVs and its application in real-time pipeline classification. Sensors 17, 1268.

Kumar, P., McElhinney, C.P., Lewis, P., McCarthy, T., 2014. Automated road markings extraction from mobile laser scanning data. Int. J. Appl. Earth Observ. Geoinform. 32, 125–137.

Liu, X., 2008. Airborne LiDAR for DEM generation: some critical issues. Progr. Phys. Geogr. 32, 31–49.

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 Sensing 10, 1531.

Mather, P., Tso, B., 2016. Classification Methods for Remotely Sensed Data. CRC Press.

Morse, B.S., 2000. Lecture 4: Thresholding. Brigham Young University.

Olsen, M.J., Parrish, C.E., Che, E., Jung, J., Greenwood, J., 2018. Lidar for Maintenance of Pavement Reflective Markings and Retroreflective Signs: Vol. I Reflective Pavement Markings, Oregan Department of Transportation.

Powers, D.M., 2007. Evaluation: from Precision, Recall and F-Factor to ROC, Informedness, Markedness and Correlation, Flinders University of South Australia, Adelaide, Australia.

Rastiveis, H., 2015. Decision level fusion of LIDAR data and aerial color imagery based on Bayesian theory for urban area classification. Int. Arch. Photogram., Remote Sensing Spatial Inform. Sci. 40, 589.

Rastiveis, H., Samadzadegan, F., Reinartz, P., 2013. A fuzzy decision making system for building damage map creation using high resolution satellite imagery. Nat. Hazards Earth Syst. Sci. 13, 455.

Rodríguez-Cuenca, B., García-Cortés, S., Ordóñez, C., 2015. An approach to detect and delineate street curbs from MLS 3D point cloud data. Automat. Construct. 51, 103–112. https://doi.org/10.1016/j.autcon.2014.12.009.

Sahoo, P.K., Soltani, S., Wong, A.K., 1988. A survey of thresholding techniques. Comput. Vision, Graphics, Image Process. 41, 233–260.

Sarasua, W.A., Clarke, D.B., Davis, W.J., 2003. Evaluation of Interstate Pavement Marking Retroreflectivity, Report to South Carolina Department of Transportation, Report Number FHWA-SC-03-01.

Serna, A., Marcotegui, B., Goulette, F., Deschaud, J.-E., 2014. Paris-Rue-Madame Database: A 3D Mobile Laser Scanner Dataset for Benchmarking Urban Detection, Segmentation and Classification Methods. In: 4th International Conference on Pattern Recognition, Applications and Methods ICPRAM 2014, ACM, Angers, France, pp. 819–824.

Shams, A., Sarasua, W.A., Famili, A., Davis, W.J., Ogle, J.H., Cassule, L., Mammadrahimli, A., 2018. Highway Cross-Slope Measurement Using Mobile LiDAR:. https://doi.org/10.1177/0361198118756371.

Smadja, L., Ninot, J., Gavrilovic, T., 2010. Road extraction and environment interpretation from Lidar sensors. IAPRS 38, 1.

Soilán, M., Riveiro, B., Martínez-Sánchez, J., Arias, P., 2017. Segmentation and classification of road markings using MLS data. ISPRS J. Photogram. Remote Sensing 123, 94–103.

Soilán, M., Riveiro, B., Sánchez-Rodríguez, A., González-deSantos, L., 2018. Application of MLS data to the assessment of safety-related features in the surrounding area of automatically detected pedestrian crossings. Int. Arch. Photogram., Remote Sensing Spatial Inform. Sci.

Thamizharasan, A., Sarasua, W.A., Clarke, D.B., Davis, W.J., 2003. A methodology for estimating the lifecycle of interstate highway pavement marking retroreflectivity. In: 82nd Annual Meeting of the Transportation Research Board of the National Academies, Washington, DC.

Vosselman, G., 2009. Advanced Point Cloud Processing. Photogrammetric Week.

University of Stuttgart Stuttgart, Germany, pp. 137–146.
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 Sensing 128, 111–129.
Wen, C., Sun, X., Li, J., Wang, C., Guo, Y., Habib, A., 2019. A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds. ISPRS J. Photogram. Remote Sensing 147, 178–192.
Yan, L., Liu, H., Tan, J., Li, Z., Xie, H., Chen, C., 2016. Scan line based road marking extraction from mobile Lidar point clouds. Sensors 16, 903.
Yang, B., Dong, Z., 2013. A shape-based segmentation method for mobile laser scanning point clouds. ISPRS J. Photogram. Remote Sensing 81, 19–30.
Yang, B., Fang, L., Li, Q., Li, J., 2012. Automated extraction of road markings from mobile Lidar point clouds. Photogram. Eng. Remote Sensing 78, 331–338.
Yang, B., Liu, Y., Dong, Z., Liang, F., Li, B., Peng, X., 2017. 3D local feature BKD to extract road information from mobile laser scanning point clouds. ISPRS J. Photogram. Remote Sensing 130, 329–343.
Yuan, Y., Shaw, M.J., 1995. Induction of fuzzy decision trees. Fuzzy Sets Syst. 69, 125–139.
Zaboli, M., Rastiveis, H., Shams, A., Hosseiny, B., Sarasua, W.A., 2019. Classification of mobile terrestrial Lidar point cloud in urban area using local descriptors. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XLII-4/W18, 1117–1122.
Zadeh, L.A., 1978. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst. 1, 3–28.
Zai, D., Li, J., Guo, Y., Cheng, M., Lin, Y., Luo, H., Wang, C., 2017. 3-D road boundary extraction from mobile laser scanning data via supervoxels and graph cuts. IEEE Trans. Intelligent Transp. Syst.
Zai, D., Li, J., Guo, Y., Cheng, M., Lin, Y., Luo, H., Wang, C., 2018. 3-D road boundary extraction from mobile laser scanning data via supervoxels and graph cuts. IEEE Trans. Intelligent Transp. Syst. 19, 802–813.
Zimmerman, H., 1996. Fuzzy Set Theory and it Applications. Kluwer, Massachusetts.