Road Centerline Extraction in Complex Urban Scenes From LiDAR Data Based on Multiple Features

Xiangyun Hu, Yijing Li, Jie Shan, *Member, IEEE*, Jianqing Zhang, and Yongjun Zhang

***Abstract*—Automatic extraction of roads from images of complex urban areas is a very difficult task due to the occlusions and shadows of contextual objects, and complicated road structures. As light detection and ranging (LiDAR) data explicitly contain direct 3-D information of the urban scene and are less affected by occlusions and shadows, they are a good data source for road detection. This paper proposes to use multiple features to detect road centerlines from the remaining ground points after filtering. The main idea of our method is to effectively detect smooth geometric primitives of potential road centerlines and to separate the connected nonroad features (parking lots and bare grounds) from the roads. The method consists of three major steps, i.e., spatial clustering based on multiple features using an adaptive mean shift to detect the center points of roads, stick tensor voting to enhance the salient linear features, and a weighted Hough transform to extract the arc primitives of the road centerlines. In short, we denote our method as Mean shift, Tensor voting, Hough transform (MTH). We evaluated the method using the Vaihingen and Toronto data sets from the International Society for Photogrammetry and Remote Sensing Test Project on Urban Classification and 3-D Building Reconstruction. The completeness of the extracted road network on the Vaihingen data and the Toronto data are 81.7% and 72.3%, respectively, and the correctness are 88.4% and 89.2%, respectively, yielding the best performance compared with template matching and phase-coded disk methods.**

***Index Terms*—Feature extraction, light detection and ranging (LiDAR), pattern recognition, remote sensing, road detection.**

I. Introduction

AUTOMATIC extraction of roads from remotely sensed data has been an open problem in remote sensing. Mena [1] and Quackenbush [2] reviewed different road extraction methods and the remote sensing data used, including multi- and hyperspectral imagery, synthetic aperture radar imagery, and light detection and ranging (LiDAR) data. Compared with road extraction in rural open or suburban areas [3], [4], complex urban scenes pose a much more difficult task. There are a number of critical issues to deal with for road extraction in urban scenes from passive remote sensing imagery. First, the occlusion of the road surface by buildings and trees makes object detection a more "ill-posed" problem due to incomplete information. Second, shadows cast by tall objects increase the spectral variance and decrease the radiometric homogeneity along the roads [5]. Third, complex road patterns and contextual objects, such as buildings, trees, and vehicles on the roads, complicate the selection of spectral and spatial features used for road detection. Finally, similar radiometry between roads and other impervious objects, such as buildings and parking lots [6], may lead to their misclassification as roads [7].

Airborne LiDAR data are produced by laser ranging to obtain 3-D geographic positions of the ground. This property makes it a better data source for road extraction in urban scenes. First, it is much easier to separate buildings from roads, although their radiometry is similar. The elevation information from LiDAR data can be also used to separate grass-covered land and rough bare earth from road areas. Second, due to its relatively narrow scanning angle (typically 20°–40° [8]), LiDAR is often free of serious occlusion of the road surface. Third, as an active sensing technology, there is no shadow effect in LiDAR data. Fourth, the intensity of LiDAR points can be used as an additional useful feature for road extraction because road surfaces have similar reflectance. Finally, water often presents as no-data areas in LiDAR data due to its absorption of laser light. This allows for the easier detection of ribbon-like ground features, such as rivers and creeks.

However, compared with imagery, LiDAR data lack spectral information, which creates difficulties for reliable object recognition. Moreover, due to the irregular distribution of LiDAR points, more effort is needed to extract accurate break lines or features, such as the edges of roads. Because of these shortcomings of LiDAR data for object extraction, researchers have integrated the processing of LiDAR data with imagery [9]–[11] or existing geodatabases [12], [13] for road extraction. Although the combined use of different data sources is theoretically better than using a single source, there are still drawbacks. First, it is either costly or even impossible to obtain different types of data for many applications. It is also difficult to precisely coregister different data sources. Second, when combining different data or using object cues from multiple sources, it needs a proper

Manuscript received November 15, 2013; revised February 13, 2014; accepted March 13, 2014. Date of publication April 21, 2014. This work was supported in part by the National Basic Research Program of China under Grant 2012CB719904 and in part by the National Natural Science Foundation of China under Grant 41171292 and Grant 41322010.

X. Hu, J. Zhang, and Y. Zhang are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: huxy@whu.edu.cn; jqz@whu.edu.cn; zhangyj@whu.edu.cn).

Y. Li is with the School of Civil Engineering and Architecture, Nanchang University, Nanchang 330031, China (e-mail: ejinn@ncu.edu.cn).

J. Shan is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China and also with the Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: shanj@whu.edu.cn).

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/TGRS.2014.2312793

fusion methodology to achieve an optimal outcome. As a result, exploring LiDAR data independent from other data sources is necessary for road extraction over complex urban scenes.

Road extraction from LiDAR data has been a focus of research since the late 1990s. Road detection in forest areas was used to generate break lines to enhance the quality of a digital terrain model [14]. Due to the aforementioned advantages, LiDAR data are increasingly used to extract roads in urban environments. Height and intensity were two major features used for classification or clustering [15], [17], [18] and segmentation [16]. To separate low bare ground and parking lots from roads, the buffered clustering approach [17] and a threshold of the maximum width of roads [8] were used to improve the correctness of road extraction. To form a vector road network, road centerlines must be detected by vectorizing, such as the phase-coded disk (PCD) approach [19].

In urban environments, it is relatively easy to filter out nonroad and tall objects, including buildings and trees, with reported algorithms [20]–[22]. The remaining problems then are as follows: 1) how can we effectively separate road areas from connected nonroad features, such as parking lots, low grassland, and bare ground; 2) how can we use multiple height and intensity features to achieve the aforementioned goal; 3) how can the processing be made less parameter sensitive and more robust to "noise" (e.g., holes in the filtered data caused by cars and trees) and to variations in the road pattern and width; and 4) how can smooth geometric primitives of roads be produced in the vector form? This paper addresses these four major research questions.

The remainder of this paper is structured as follows. Section II presents an overview of the proposed method and describes the three major processing steps after filtering. These steps include spatial clustering based on multiple features using an adaptive mean shift to detect the center points of roads, stick tensor voting for enhancing salient linear features, and a weighted Hough transform for extracting the arc primitives of road centerlines. Section III uses the Vaihingen and Toronto data sets from the International Society for Photogrammetry and Remote Sensing (ISPRS) Test Project on Urban Classification and 3-D Building Reconstruction [23] to validate the method. We compare the performance of the method with that of the template matching [24] and PCD [19] algorithms. Finally, Section IV offers the concluding remarks and discusses on future improvement to our method.

II. Road Centerline Detection Based on Multiple Features

A. Overview

Fig. 1 shows the workflow of the proposed method. The initial step, i.e., filtering, is to classify the point clouds into ground and nonground classes. We adopted the multidirectional ground filtering algorithm [25] for this purpose. It combines the elevation differences and slopes of neighboring points in various directions to filter the ground points. The subsequent steps include: 1) spatial clustering using an adaptive mean shift to detect the center points of roads; 2) stick tensor voting to enhance the salient linear features; and 3) the Hough transform to detect the arc primitives of the road centerlines. In short and to reflect its key implementation steps, the proposed method is denoted by Mean shift, Tensor voting, Hough Transform (MTH).

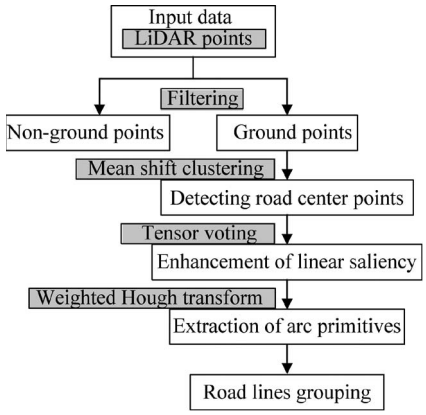


Fig. 1. Workflow of the proposed road centerline detection method.

B. Adaptive Mean Shift for Spatial Clustering to Detect the Center Points of Roads

After filtering, the remaining points are ground points, which may be roads, bare ground, parking lots, low grassland, etc. A desired property for road extraction from LiDAR points is the ability to separate nonroad areas from road areas and to make the detection robust to the existing connected nonroad areas. For this purpose, several important road features should be used, including the ribbon-like shape constraint, the smoothness of the surface, and the consistency of reflectance. This would further lead to the road center points, which are defined as points located at the center of the ribbon areas with a local smooth surface and consistent intensity. We also expect such algorithm to be insensitive to the variation in the road width and the "holes" left by the filtered cars and trees.

We now develop a spatial clustering algorithm based on a mean shift to directly distinguish the center points of the ribbon roads from ground points without rasterization. The mean shift procedure was proposed by Fukunaga and Hostetler [26] and has been widely used in image processing and pattern recognition [27]. When a uniform kernel is used, each point x iteratively shifts to the average $m(x)$ of all points in its neighborhood until a stable $m(x)$ is reached. $m(x)$ is given by

$$m(x) = \frac{1}{n} \sum_{x_i \in N(x)} x_i \tag{1}$$

where $N(x)$ is a set of neighbor points of x, and n is the number of points within $N(x)$.

Fig. 2 illustrates how the mean shift works to find the center points of the road (ribbon). With the mean shift, the points converge to the mode point (pg), as shown in Fig. 2(a). If the clustering window size is equal to the width of the road, all the points spatially distributed as a ribbon-like road are clustered to the centerline, as shown in Fig. 2(b) and (c). At the same time, the nonroad points, such as vegetation points with different features, will be separated, as shown in Fig. 2(d). Fig. 2(e) illustrates the shift vectors formed by connecting the initial point (pe) and its final position (pg) after the shifting.

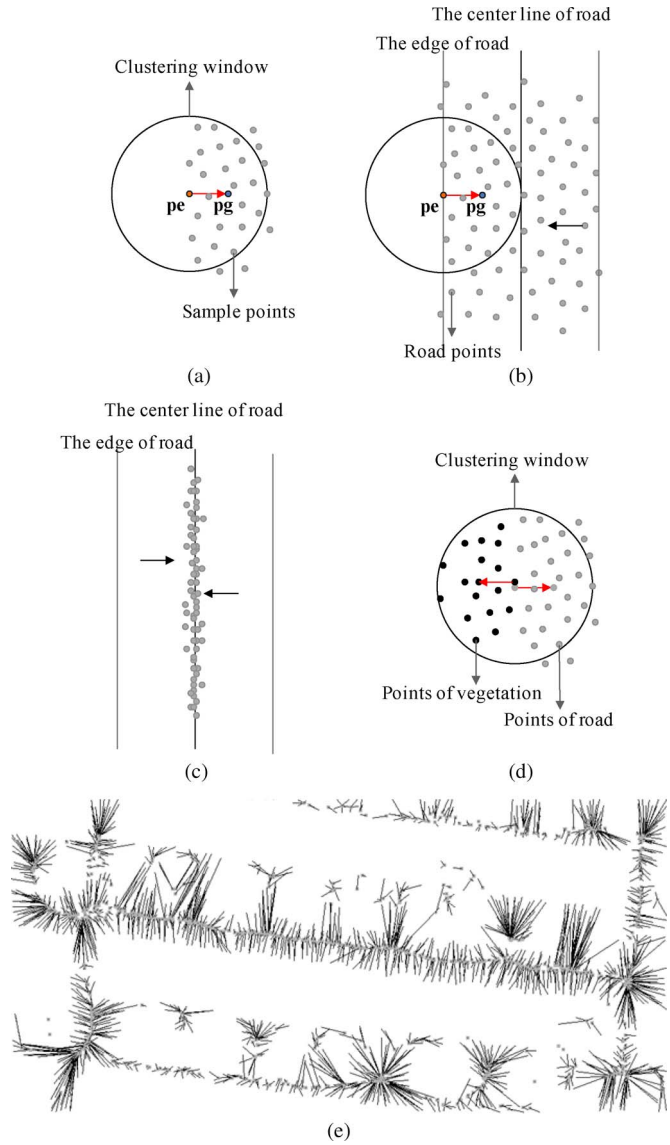


Fig. 2. Detection of road center points by the mean shift. (a) Mean shift clustering. (b) Clustering points into the road centerline. (c) Road points after clustering. (d) Separating the nonroad points from the ground points by converging them into their mode point based on distinct features, e.g., reflectance intensity and surface roughness. (e) Shift vectors shown as line segments.

Fig. 3. Mean shift clustering for the detection of road center points with holes in the area. (a) Road points with holes. (b) Clustered mode points by the mean shift.

Fig. 3(a) shows the road points with several holes left by the filtered vehicles. Fig. 3(b) is the clustering result by the mean shift with the window size set as the road width. This demonstrates that the mean shift is an appropriate approach for detecting the center points of ribbon roads, even when roads are occulted by cars and trees.

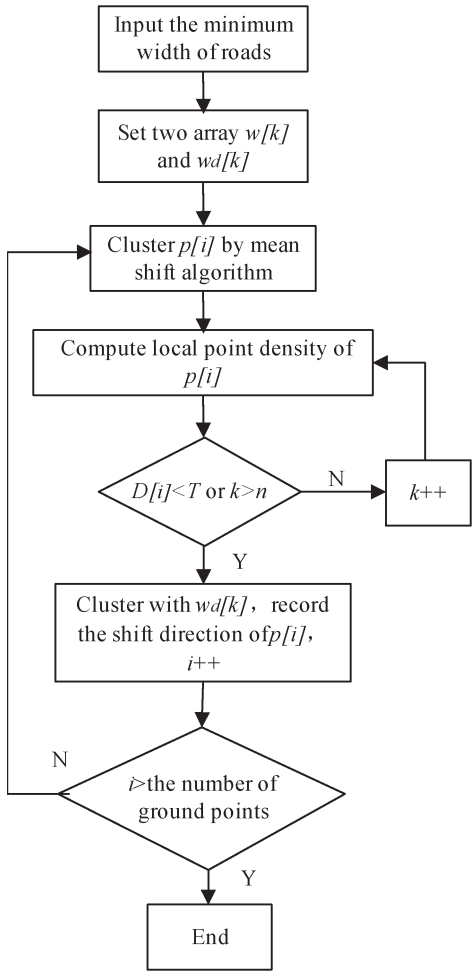


Fig. 4. Road center point detection with an adaptive window size in mean shift clustering.

To utilize the intensity and the local surface roughness to separate nonroad features, a feature vector of each point is constructed as $P_i\ (x_i,\ y_i,\ z_i,\ I_i,\ S_i)$, where $(x_i,\ y_i,\ z_i)$ are the coordinates of a LiDAR point P_i, I_i is its intensity, and S_i is the smoothness of the local surface at P_i, i.e.,

$$S_i = \sum_{j=1}^{n} (z_{ij} - \overline{z}_i)/n \tag{2}$$

where z_{ij} is the elevation of a point within the clustering window, and $\overline{z}_i$ is the average elevation of all the points in this window.

As the road width may vary, we propose mean shift clustering with an adaptive window size. As shown in Fig. 4, it can be summarized in three main steps.

Step 1) Set two arrays $w[k]$ and $w_d[k], k=1, 2, 3, \ldots, n$ ($n < 5$ will satisfy the need for most of the data); $w[k]$ stores the window sizes used to calculate the local point density, which is defined as $w[k] = cr_w$, $c = 2k + 1$; and $w_d[k]$ stores the window size used to cluster points, which is given by $w_d[k] = kr_w$, where r_w is the minimum road width.

Step 2) Cluster the ground points one by one with the adaptive window size. The uniform kernel is used in the mean

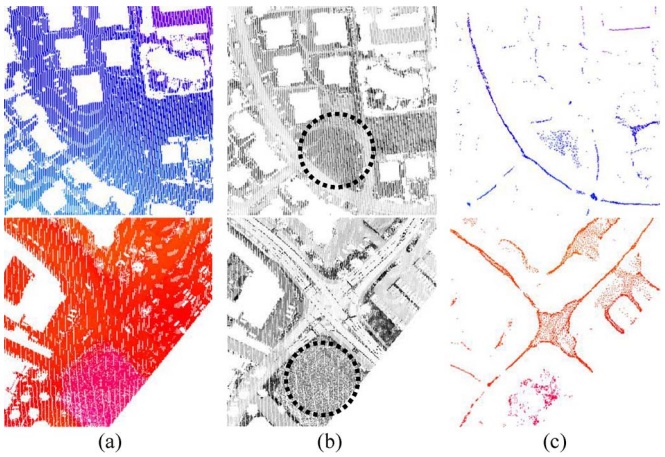


Fig. 5. Two local regions in the Vaihingen data. (a) Ground points color coded by elevation. (b) Intensity image of the ground points. (c) Results of the road center point detection.

shift procedure. Calculate the local density of point i, which is defined as $D[i] = n_i/(w[k] \times w[k])$, where n_i is the number of points in the window with size $w[k]$. Compare $D[i]$ and T, and if $D[i] < T$ or $k > n$, cluster point i with window size $w_d[k]$, and then, record the new coordinates and shift the direction of point i; otherwise, recalculate $D[i]$ by $w[k+1]$. T is the threshold of density, which is given by $T = k/c$.

Step 3) Continue the aforementioned process until all the ground points are processed.

The advantages of this algorithm are as follows. First, it provides a data-driven method that does not require *a prior* road model and is suitable for all types of ribbon roads. Second, it is insensitive to the holes left by vehicles and trees on roads. Third, it is straightforward to use multiple features, such as both intensity and surface smoothness, to distinguish road and nonroad points. To demonstrate this, Fig. 5 presents two examples of the road center point detection results from the ISPRS Vaihingen data set [23]. Here, large areas of bare ground or grass (as shown in the circled areas) become scattered points. These points can be removed by further detection of salient linear structures (the road centerlines). Fig. 6 shows the clustering results for the entire test data sets of the Vaihingen and Toronto areas.

C. Stick Tensor Voting to Enhance Linear Saliency

With the mean shift, the ribbon road points in the LiDAR data are clustered to their center locations, demonstrating the linear patterns from which the road centerlines can be extracted. However, as shown in Fig. 5, there are still large numbers of scattered points of parking lots and grass areas, which interfere with the subsequent extraction of road centerlines. To detect the correct road center points, we use tensor voting [28] to enhance their linear saliency and reduce the saliency of the nonroad scattering points.

An important advantage of using tensor voting is its ability to detect linear features from the data points with only one parameter, i.e., the size of voting fields. Tensor voting is a powerful method of perceptual grouping, which comprises the following three steps.

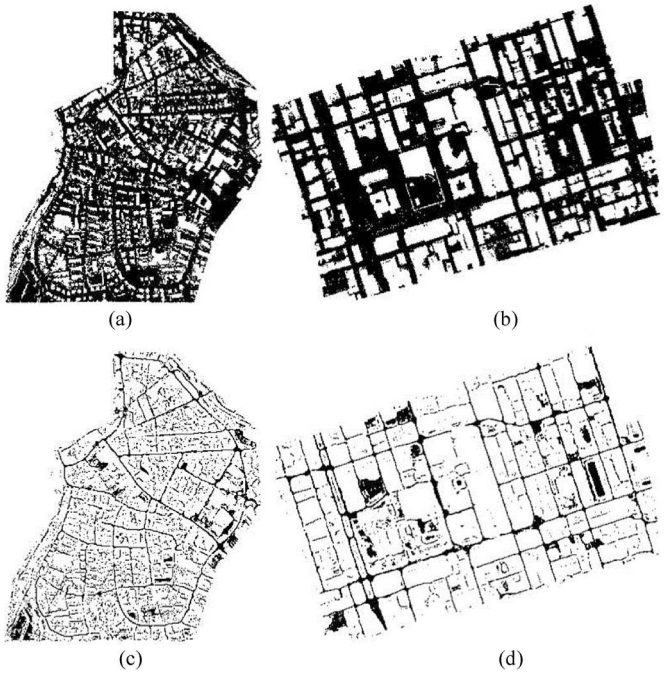


Fig. 6. Detection of road center points for the two test data sets. The ground points (black) in the (a) Vaihingen data and the (b) Toronto data. The detected road center points for the (c) Vaihingen data and the (d) Toronto data.

1) A point with a tangent or normal direction, or no direction, is expressed in the form of a tensor. Assuming the points are in a 2-D space, the tensor is computed by

$$T = (\lambda_1 - \lambda_2)\vec{e}_1\vec{e}_1^T + \lambda_2(\vec{e}_1\vec{e}_1^T + \vec{e}_2\vec{e}_2^T) = \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix} \quad (3)$$

where λ_1 and λ_2 $(\lambda_1 \geqslant \lambda_2)$ are the eigenvalues, $\vec{e}_1$ and $\vec{e}_2$ are the corresponding eigenvectors, and θ is the normal direction computed by the shift vector at the mean shift stage. $\vec{e}_1\vec{e}_1^T$ describes a stick, whereas $(\vec{e}_1\vec{e}_1^T + \vec{e}_2\vec{e}_2^T)$ describes a plate. If $\lambda_1 = 1$ and $\lambda_2 = 0$, T is a stick tensor with strong directionality, which is beneficial to enhance the linear saliency. This way, we treat each point as a virtual stick that has a normal direction for tensor voting.

2) Each input site (tensor T) collects all the votes from its neighborhood points through a predefined tensor field and, finally, changes into a new tensor. The tensor field used here is the same as that in [28]. The size of this field is a function of the scale of voting σ, as described in [28].

3) The linear saliency factor S_F of each point is calculated by

$$S_F = \lambda_1 - \lambda_2 \quad (4)$$

where S_F represents the saliency of the linear feature passing through a point. The greater the value, the more likely that the point is a road point and not a point of a parking lot, as shown in Fig. 7.

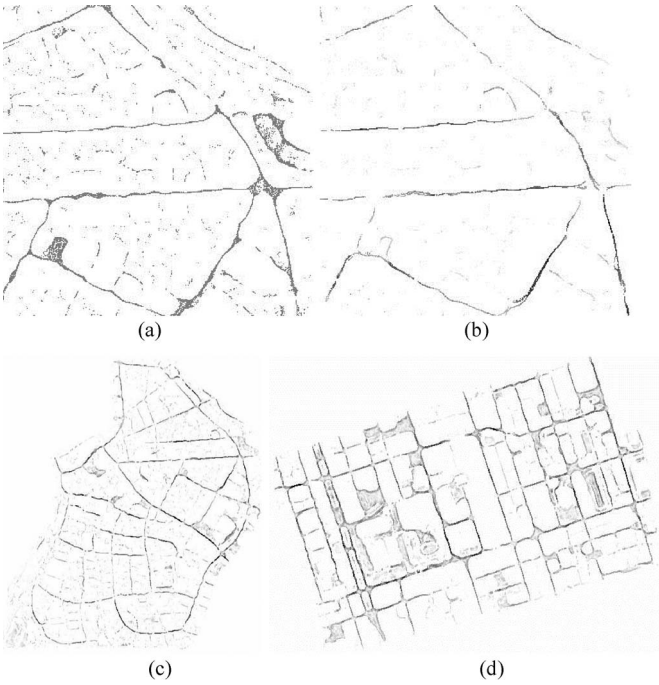


Fig. 7. Saliency images after stick tensor voting, in which the darker pixels indicate a larger S_F. (a) Clustered road center points by the mean shift. (b) Linear saliency of (a) after stick tensor voting. (c) Saliency image of the Vaihingen data. (d) Saliency image of the Toronto data.

D. Weighted Hough Transform for the Extraction of Arc Primitives

We use arcs to represent the road primitives for two reasons: 1) arcs are suitable for delineating the smooth parts of a road centerline, and connecting the subsequent arcs makes a complete road map; and 2) arcs are suitable for a grouping algorithm to form a road network, whether the road is straight or curved. A straight road could be considered an arc with an infinite radius, whereas a curved road may be composed of several arcs with different radii.

The Hough transform [29] is a method for estimating the parameters of a shape from its discrete points by voting and is relatively less affected by noise. The technique we used groups the points into the vector arc primitives of roads by considering the linear saliency of a point in the voting process and, thus, improves the reliability of the extraction results. The weight of each point for voting is given by

$$w_e = \frac{S_F}{\text{Max}(S_F)} \tag{5}$$

where $\text{Max}(S_F)$ is the global maxima of the saliency factors obtained from the tensor voting step. There are four main iterative steps in the arc extraction as follows.

1) Find point P_k with the current maximal saliency factor $\text{Max}(S_F)$. For a given saliency threshold Ts, if $\text{Max}(S_F) > \text{Ts}$, go to step 2; otherwise, the algorithm terminates, and all the points are ignored.
2) Extract a circle passing through point P_k by the Hough transform. According to the equation of a circle, i.e., $(x - x_0)^2 + (y - y_0)^2 = r^2$, there are three unknown parameters, i.e., the circle's center (x_0, y_0) and radius r. However, point P_k is considered a point on the circle; therefore, only two parameters need to be calculated to determine this circle, i.e., radius r and the direction angle θ between point P_k and the center of the circle, as illustrated in Fig. 8. The two parameters are obtained at the local maxima in the accumulator space, where $r \in (50\text{ m}, 10000\text{ m})$, and the steps of r and θ are 50 m and 2°, respectively.
3) Achieve the two endpoints of the arc primitives on the detected circle. The process is implemented by the grouping points that are close to the circle, starting from P_k with a small gap. Then, remove these points to reduce the interference to the detection of the remaining arcs (see Fig. 9).
4) Continue the aforementioned process until all the points are processed.

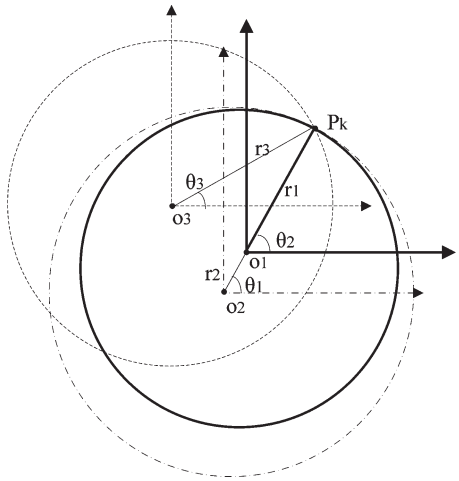


Fig. 8. Two parameters (θ, r) of a circle passing a known point P_k.

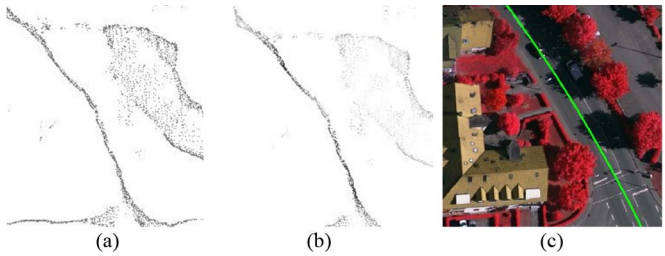


Fig. 9. Example of the detection of arc primitives. (a) Clustered center points. (b) Saliency image by tensor voting. (c) Detected arc by the Hough transform.

III. Experiments and Evaluation

An experimental study has been carried out in order to validate our method. The Vaihingen and Toronto data were provided by the ISPRS Test Project on Urban Classification and 3-D Building Reconstruction [23]. The LiDAR data for Vaihingen were captured by a Leica ALS50 system, and the point density is 4 points/m^2. The LiDAR data for Toronto were acquired by Optech, Inc., and the point density is approximately 6 points/m^2 [23]. Vaihingen is a typical European town, and Toronto is a developed North American city with dense high-rise buildings. Accordingly, these data sets provide a good sample to test the feature extraction methods in complex urban

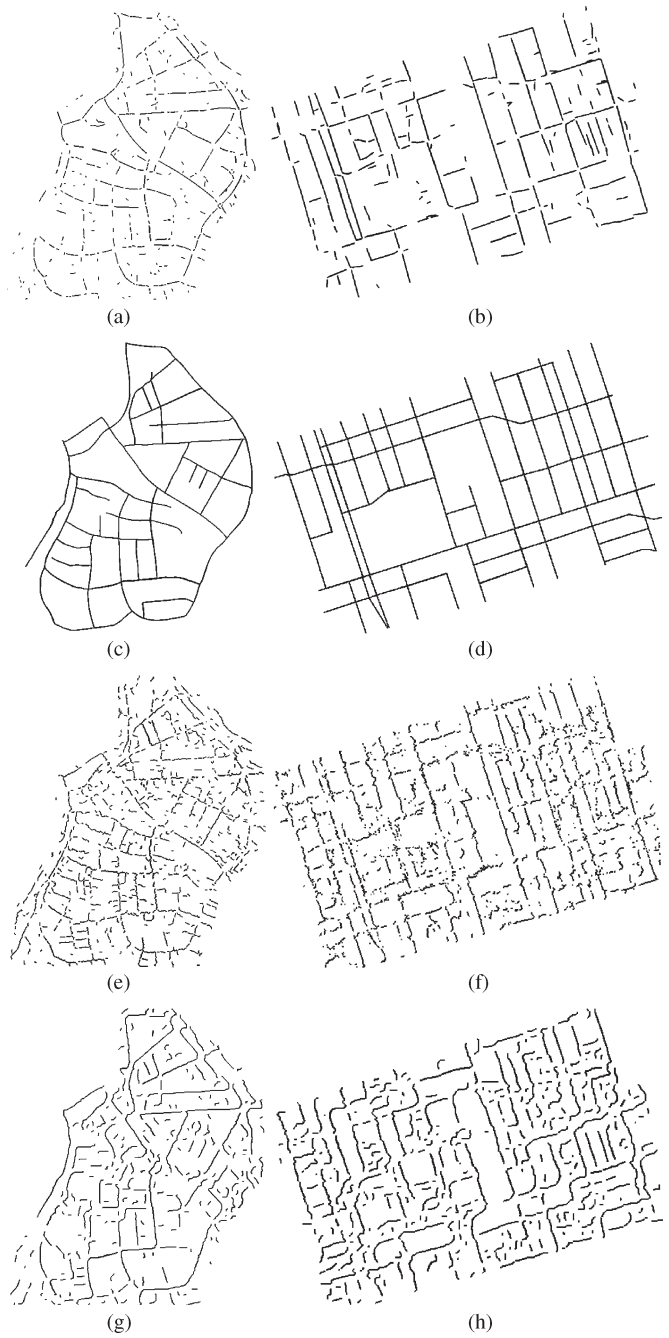


Fig. 10. Extracted road primitives for (left) Vaihingen and (right) Toronto. (a) and (b) Detected road primitives by our method. (c) and (d) Reference road map digitized manually. (e) and (f) Primitives detected by template matching. (g) and (h) Primitives detected by the PCD method.

scenes. The two data sets contain various types of roads with different widths and many other man-made and natural objects, such as buildings, trees, parking lots, and grassland. Much of the road surface is covered by vehicles and the shadows of surrounding tall objects.

The arc primitives of the roads extracted by our method are shown in Fig. 10(a) and (b). Fig. 10(c) and (d) presents the road maps that were manually digitized from the data. We used these road maps as references to evaluate the extraction results, and we compared our method with two other methods of road detection, i.e., template matching [24] and the PCD approach [19].

TABLE I
PARAMETERS OF THE PROPOSED METHOD

	Vaihingen	Toronto
Minimal width of roads(meters)	5	7
Clustering window size of intensity	30	30
Clustering window size of elevation(meters)	0.5	0.5
Clustering window size of smoothness(meters)	0.05	0.05
The scale of voting σ	9	9
The threshold of saliency factor S	200	200

TABLE II
PARAMETERS OF THE TEMPLATE MATCHING METHOD

	Vaihingen	Toronto
Average width of roads (pixels)	20	25
Minimal width of roads (pixels)	10	10
Maximal width of roads (pixels)	25	30
Average gray of roads	200	200
Smooth scale	3.0	3.0
Scale of morphological dilating (pixels)	2	2

TABLE III
PARAMETERS OF THE PCD ALGORITHM

	Vaihingen	Toronto
Intensity threshold	70	30
Radius of phase disk(pixels)	25	10
Scale of morphological closing (pixels)	2	2

Template matching is a method of detecting road center points based on the analysis of the road profile [24]. To carry out this approach, raster images first need to be generated. An intensity image of the filtered ground points was created for template matching. The pixel size is selected as 0.8 m, and its value is determined by

$$g_i = \begin{cases} 255 - I_i \frac{255}{I_{\max} - I_{\min}}, & i - \text{ground points} \\ 0, & i - \text{nonground points} \end{cases} \quad (6)$$

where $I_{\max}$ and $I_{\min}$ are the global maximum and minimum intensities of the ground points. Fig. 10(e) and (f) show the extraction results by template matching.

For the PCD algorithm, the same raster image is first created. An intensity threshold is then used to separate the road points from the ground points. The resultant binary image with a 0.75-m pixel size is convolved with a PCD, yielding two images, i.e., a magnitude image and a phase image. Road centerlines are extracted by tracing the ridge of the magnitude image, and the direction of the lines can be retrieved from the phase image [19]. The extraction results are shown in Fig. 10(g) and (h).

All the methods need to set a number of parameters. They are chosen by evaluating the data over the test areas and by trial and error. The final selections for the three methods are listed in Tables I–III, respectively. In our method, the minimal width of roads is set as 5 and 7 m, respectively, for Vaihingen and Toronto. The clustering window size (30, unitless) of intensity is used to separate the road points from other objects with

TABLE IV
QUALITY OF THE THREE METHODS FOR ROAD PRIMITIVE EXTRACTION

Quality metrics	MTH		TM		PCD	
	Vaih.	Toro.	Vaih.	Toro.	Vaih.	Toro.
True Pos.(m)	5203.6	12162.6	6050.1	7112.3	5684.5	9322.4
False Pos.(m)	1844.2	5052.9	6164.3	7665.6	5005.0	9112.8
False Neg.(m)	4543.5	4291.2	3697.0	9341.6	4062.6	7131.5
Completeness %	53.4	73.9	62.1	43.2	58.3	56.7
Correctness %	73.8	70.6	49.5	48.1	53.2	50.6
Overall Quality%	44.9	56.5	37.9	29.5	38.5	36.5

different materials, particularly vegetation, which is chosen to cluster points with homogenous intensity and tolerate the noise of intensity. The clustering window size of elevation is chosen as 0.5 m considering the typical slope of urban roads, whereas the smoothness tolerance is set as 0.05 m. The scale of voting σ corresponds to the size of the tensor field. A smaller scale is beneficial to preserve details, whereas a larger scale is robust to noise and beneficial to bridge gaps but at a heavier computational cost. The threshold of saliency factor S is determined based on the saliency factors of the nonroad points. As for the parameters of the other two methods, they were set by trial and error or equivalently set to the parameters of our method, e.g., the minimal width in meters or in pixels. To make a fair comparison, we tuned the parameters to produce the best results.

A number of observations can be noted when evaluating the results in Fig. 10. The primitives extracted by our method are more connected or less fragmented than those of the other two methods. This is likely due to the use of the Hough transform that utilizes the embedded piecewise circular assumption and is therefore able to bridge the small gaps between the extracted road center points along the road direction. Another observation to note is that our approach yields "cleaner" results than the other two methods in the sense that there are less isolated and small road primitives. This is the effect of the tensor voting step, where scale parameter σ is used to control the level of details of the extracted road primitives. Our study has tuned this parameter according to the nature of the scene and the resolution of the input data with the focus on arteries in urban scenes. Finally, our extracted road primitives are apparently smoother than the primitives from the other two methods. This is not only aesthetic but more realistic for an urban area. Such an effect is contributed by the use of tensor voting to enhance the linear resilience and the subsequent Hough transform that accommodates road primitives of varying curvatures and lengths.

Table IV summarizes the quality metrics [30] for the road primitives extracted by the three methods. The evaluation is referenced to the manually digitized roads. The true positive and the false positive are, respectively, the correctly and wrongly extracted roads, whereas the false negative refers to the missed roads, all measured in meters. The relative quality metrics are defined as

$$\begin{aligned} \text{correctness} &= \frac{\text{TP}}{\text{TP}+\text{FP}} \times 100\% \\ \text{completeness} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \times 100\% \\ \text{quality} &= \frac{\text{TP}}{\text{TP}+\text{FP}+\text{FN}} \times 100\%. \end{aligned} \tag{7}$$

(a) (b)

Fig. 11. Extracted main road networks of the (a) Vaihingen data and a (b) part of the Toronto data.

Notably, our method achieves the best qualities among the three methods. This is more apparent when looking at a correctness of over 70% for both data sets. As a result of introducing tensor voting and the Hough transform, they can help select points on the roads and fit them to a predefined road model (a straight line or a circle). It is interesting to note that the completeness (53.4%) of our method for the Vaihingen data set is not as good as those of the other two methods. Since many intersections in the data set are so close to each other, the road segments between intersections with no salient linear features are ignored. However, the length of the wrongly extracted roads extracted by our method is significantly less than those extracted by the other two methods, which is very beneficial to road networks grouping.

Finally, the root-mean-square error of the results [30] for the Vaihingen data and the Toronto data are, respectively, 1.16 and 1.25 m. This is as expected since it is about the point ground spacing of the two input LiDAR data sets, i.e., $\text{sqrt}(4)/2 = 1.0$ m (for Vaihingen) and $\text{sqrt}(6)/2 = 1.2$ m (for Toronto). This level of accuracy is similar to an early study for building extraction and regulation by Sampath and Shan [31]. This demonstrates that our effort reaches the best accuracy the input data sets can provide.

The final road networks are formed by connecting the extracted road primitives. A hierarchical grouping strategy [32] was adopted for this purpose. In this process, the gaps between the primitives are bridged, and the isolated primitives are grouped into complete and connected roads. Short lines that are likely not meaningful roads are removed. Fig. 11 demonstrates the extracted networks of main roads with our approach, whereas Table V lists the quality metrics for all three methods. The quality of the extracted main road networks exhibits a noticeable improvement compared with Table IV for road primitives. This demonstrates the necessity of the final grouping

TABLE V
QUALITY OF THE THREE METHODS FOR ROAD EXTRACTION AFTER PRIMITIVE GROUPING

Quality metrics	MTH		TM		PCD	
	Vaih.	Toro.	Vaih.	Toro.	Vaih.	Toro.
True Pos.(m)	7960.5	11898.2	7020.5	11194.5	6508.5	10239
False Pos.(m)	1040.6	1442.7	2707.5	2769	1475.4	4729.5
False Neg.(m)	1786.6	4555.7	2726.5	5259.3	3238.6	6214.8
Completeness %	81.7	72.3	72.0	68.0	66.8	62.2
Correctness %	88.4	89.2	72.2	80.2	81.5	68.4
Overall Quality%	73.8	66.5	56.4	58.2	58.0	48.3

for road primitives. Such positive effect is observed for all three methods. A reliable road extraction method should make an effort in bridging the gaps between road primitives, removing false alarms, and cleaning its results. Moreover, our method apparently presents a better outcome than template matching and the PCD method for both data sets. The completeness of the road network extraction on the Vaihingen data and the Toronto data are 81.7% and 72.3%, respectively, and the correctness are 88.4% and 89.2%, respectively. This, in turn, implies that finding quality road primitives is an important prerequisite for a successful road network extractor, and the proposed method did contribute to this end.

IV. DISCUSSION AND CONCLUSION

This paper has proposed a multistep and multifeature-based method (the MTH) for the automatic detection of roads from the LiDAR data in complex urban scenes. The initial step is to classify the LiDAR points as ground and nonground classes. Three steps are then carried out on the ground points to detect the arc primitives of roads. First, the center points of roads are detected using an adaptive mean shift. Second, the salient linear features are enhanced using stick tensor voting. Finally, the arc primitives of the road centerlines are extracted using a weighted Hough transform. The major contribution of this paper is that the proposed sequential processing method makes full use of the salient features of the roads in the LiDAR data on urban scenes. Specifically, the mean shift algorithm integrates the surface smoothness, the ribbon-like shape, and the consistent reflectance of the roads into a data-driven model; the shift vector is then used to enhance the linear saliency by tensor voting, in which a point is treated as a virtual stick with direction; and finally, the saliency factors are used in a weighted Hough transform to iteratively detect the salient arc primitives of roads. The experimental results in the ISPRS test data sets show that our method can effectively detect roads in complex urban scenes with varying road widths and noise, such as the holes left by cars on the road. As a result, it produces better road primitives than the template matching and PCD algorithms. It is shown that good road primitives are necessary for generating a complete road network. Our method can provide a quality input for the subsequent hierarchical grouping for the final road network generation.

As the experiments revealed, there are problems in our method that still need to be resolved. One problem is the heavy computational cost in the tensor voting step. For the Vaihingen data, the template matching and PCD algorithms take 1.5 and 1.2 min, respectively, whereas our method needs 97 min. The computer we used is a personal computer with a 2-GHz central processing unit. However, due to the graphic-processing-unit computing, the highly parallel local computing of tensor voting and the mean shift can hopefully be greatly accelerated. The other problem is the recognition of the contextual objects of roads, such as lane markings, road junction patterns, vehicles, and road edges. High-resolution imagery can certainly provide useful cues to diagnose the extracted road primitives and improve the quality of the extraction. However, there are still many issues in the integration of high-resolution images with the LiDAR points for road extraction that need to be addressed. For instance, how can multiple cues be detected from the complex images? Furthermore, how should we handle cues that have very complex semantic and spatial relations? Our experiments suggest that a feasible way might be to first extract the major salient roads from the LiDAR points and, then, use images to complete and correct the initial results. This aims to form the basis of our future work to ultimately develop an operational system of automatic road extraction for complex urban scenes.

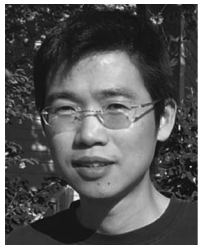
ACKNOWLEDGMENT

The authors would like to thank the German Society for Photogrammetry, Remote Sensing and Geoinformation [23] http://www.ifp.unistuttgart.de/dgpf/DKEP-Allg.html for providing the Vaihingen data set and Optech, Inc.; First Base Solutions, Inc.; York University, Toronto, ON, Canada; and the International Society for Photogrammetry and Remote Sensing Working Group III/4 for providing the Toronto data set.

REFERENCES

[1] J. B. Mena, "State of the art on automatic road extraction for GIS update: A novel classification," *Pattern Recognit. Lett.*, vol. 24, no. 16, pp. 3037–3058, Dec. 2003.
[2] L. J. Quackenbush, "A review of techniques for extracting linear features from imagery," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 12, pp. 1383–1392, Dec. 2004.
[3] A. Baumgartner, C. Steger, H. Mayer, W. Eckstein, and H. Ebner, "Automatic road extraction based on multi-scale, grouping, and context," *Photogramm. Eng. Remote Sens.*, vol. 65, no. 7, pp. 777–785, Jul. 1999.
[4] C. S. Zhang, "Towards an operational system for automated updating of road databases by integration of imagery and geodata," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 3/4, pp. 166–186, Jan. 2004.
[5] Q. H. Weng, "Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends," *Remote Sens. Environ.*, vol. 117, no. 15, pp. 34–49, Feb. 2012.
[6] R. Péteri and T. Ranchin, "Road networks derived from high spatial resolution satellite remote sensing data," in *Remote Sensing of Impervious Surfaces*, Q. Weng, Ed. New York, NY, USA: Taylor & Francis, 2007, pp. 215–236.
[7] R. Gecen and G. Sarp, "Road detection from high and low resolution satellite images," in *Proc. ISPRS Archives*, Beijing, China, 2008, pp. 355–358.
[8] F. Rottensteiner and S. Clode, "Building and road extraction by LiDAR and imagery," in *Topographic Laser Ranging and Scanning: Principles*

and Processing., 1st ed. Boca Raton, FL, USA: CRC Press, 2008, pp. 445–478.
[9] X. Hu, C. V. Tao, and Y. Hu, "Automatic road extraction from dense urban area by integrated processing of high resolution imagery and LiDAR data," in *Proc. ISPRS Archives*, Istanbul, Turkey, 2004, pp. 388–400.
[10] P. Zhu, Z. Lu, X. Chen, K. Honda, and A. Elumnoh, "Extraction of city roads through shadow path reconstruction using laser data," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 12, pp. 1433–1440, Dec. 2004.
[11] G. Wang, Y. Zhang, J. Li, and P. Song, "3D road information extraction from LiDAR data fused with aerial-images," in *Proc. IEEE ICSDM*, Fuzhou, China, 2011, pp. 362–366.
[12] C. Hatger and C. Brenner, "Extraction of road geometry parameters from laser scanning and existing databases," in *Proc. ISPRS Archives*, Dresden, Germany, 2003, pp. 225–230.
[13] A. Boyko and T. Funkhouser, "Extracting roads from dense point clouds in large scale urban environment," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 6, pp. 2–12, Dec. 2011.
[14] W. Rieger, M. Kerschner, T. Reiter, and F. Rottensteiner, "Roads and buildings from laser scanner data within a forest enterprise," in *Proc. Int. Archives Photogramm. Remote Sens.*, Washington, DC, USA, 1999, pp. 642–649.
[15] S. Clode, P. J. Kootsookos, and F. Rottensteiner, "The automatic extraction of roads from LiDAR data," in *Proc. ISPRS*, Istanbul, Turkey, 2004, pp. 231–236.
[16] S. Clode, F. Rottensteiner, and P. J. Kootsookos, "Improving city model determination by using road detection from LIDAR data," in *Proc. DAGM CMRT*, Vienna, Austria, 2005, pp. 159–164.
[17] Y. W. Choi, Y. W. Jang, H. J. Lee, and G. S. Cho, "Three-dimensional LiDAR data classifying to extract road point in urban area," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 725–729, Oct. 2008.
[18] F. Samadzadegan, M. Hahn, and B. Bigdeli, "Automatic road extraction from LIDAR data based on classifier fusion," in *Proc. ISPRS Archives*, Shanghai, China, 2009, pp. 1–6.
[19] S. Clode, F. Rottensteiner, P. Kootsookos, and E. Zelniker, "Detection and vectorisation of roads from LIDAR data," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 5, pp. 517–535, May 2007.
[20] H. Masaharu and K. Ohtsubo, "A filtering method of airborne laser scanner data for complex terrain," in *Proc. ISPRS Archives*, Graz, Austria, 2002, pp. 165–169.
[21] K. Zhang, S. Chen, D. Whitman, M. Shyu, J. Yan, and C. Zhang, "A progressive morphological filter for removing nonground measurements from airborne LIDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 4, pp. 872–882, Apr. 2003.
[22] G. Sithole and G. Vosselman, "Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 59, no. 1/2, pp. 85–101, Aug. 2004.
[23] M. Cramer, "The DGPF test on digital aerial camera evaluation—Overview and test design," *Photogramm.-Fernerkundung-Geoinf.*, vol. 2, pp. 73–82, Jan. 2010.
[24] X. Hu and C. V. Tao, "A reliable and fast ribbon road detector using profile analysis and model-based verification," *Int. J. Remote Sens.*, vol. 26, no. 5, pp. 887–902, Mar. 2005.
[25] X. Meng, L. Wang, J. L. Silvan-cardenas, and N. Currit, "A multi-directional ground filtering algorithm for airborne LiDAR," *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 1, pp. 117–124, Jan. 2009.
[26] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 32–40, Jan. 1975.
[27] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
[28] P. Mordohai and G. Medioni, "Tensor voting," in *Tensor Voting: A Perceptual Organization Approach to Computer Vision and Machine Learning*., 1st ed. San Rafael, CA, USA: Morgan & Claypool, 2006, pp. 12–19.
[29] P. Hough, "Method and means for recognizing complex patterns," U.S. Patent 3 069 654, Dec. 18, 1962.
[30] C. Wiedemman, C. Heipke, H. Mayer, and O. Jamet, "Empirical evaluation of automatically extracted road axes," in *Empirical Evaluation Methods in Computer Vision*. Piscataway, NJ, USA: IEEE Press, 1998, pp. 172–187.
[31] A. Sampath and J. Shan, "Building boundary tracing and regularization from airborne LIDAR point clouds," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 7, pp. 805–812, Jul. 2007.
[32] X. Hu and C. V. Tao, "Automatic extraction of main road centerlines from high resolution satellite imagery using hierarchical grouping," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 9, pp. 1049–1056, Sep. 2007.



Xiangyun Hu received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2001.

From 2002 to 2005, he was a Postdoctoral Research Fellow with the Department of Earth and Space Science and Engineering, Lassonde School of Engineering, York University, Toronto, ON, Canada. He developed a semiautomatic feature extraction technology "SmartDigitizer," which was acquired by PCI Geomatics, Leica Geosystems, and Microsoft. From 2005 to 2010, he was a Senior Software Engineer with ERDAS, Inc., Atlanta, GA, USA. He is currently a Professor with the School of Remote Sensing and Information Engineering, Wuhan University. He is the author or coauthor of more than 40 papers published in journals and conferences in intelligent analysis and feature extraction of remotely sensed data.

Yijing Li received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2013.

She is currently with the School of Civil Engineering and Architecture, Nanchang University, Nanchang, China. Her areas of interests include pattern recognition and object extraction from images and light detection and ranging (LiDAR) data.

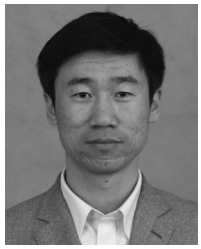
Jie Shan (M'08) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China.

He has held faculty positions with universities in China and Sweden, and he has been a Research Fellow in Germany. He is with the Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, USA, and also with the School of Remote Sensing and Information Engineering, Wuhan University. His areas of interests include sensor geometry, pattern recognition from images and light detection and ranging (LiDAR) data, object extraction and reconstruction, urban remote sensing, and automated mapping.

Dr. Shan has been the Cochair of the Remote Sensing Data Fusion Working Group and the Pattern Recognition for Remote Sensing Working Group, both in the International Society for Photogrammetry and Remote Sensing. He is an Associate Editor for the IEEE TRANSACTIONS GEOSCIENCE AND REMOTE SENSING and an Assistant Editor for *Photogrammetric Engineering and Remote Sensing*. He was a recipient of multiple academic awards, among which are the Talbert Abrams Grand Award and the Environmental Systems Research Institute Award for Best Scientific Paper in Geographic Information Systems (First Place).

Jianqing Zhang received the B.S. degree in mathematics from Nankai University, Tianjin, China, in 1969 and the M.S. degree from Wuhan University, Wuhan, China, in 1982.

He is currently a Professor with the School of Remote Sensing and Information Engineering, Wuhan University. He is the author or coauthor of more than 150 papers. His research interests include automation in photogrammetry, 3-D computer vision for city modeling, and the processing of remote sensing images.



Yongjun Zhang received the B.S., M.S., and Ph.D. degrees from Wuhan University, Wuhan, China, in 1997, 2000, and 2002, respectively.

He is currently a Professor of photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, Wuhan University. His research interests include space, aerial, and low-attitude photogrammetry; image matching; combined bundle adjustment with multisource data sets; 3-D city reconstruction; and industrial inspection.