Extraction of Dense Urban Buildings From Photogrammetric and LiDAR Point Clouds

LIANG GUO, XINGDONG DENG[ID]**, YANG LIU**[ID]**, HUAGUI HE, HONG LIN, GUANGXIN QIU, AND WEIJUN YANG**
Guangzhou Urban Planning and Design Survey Research Institute, Guangzhou 510060, China
Guangdong Enterprise Key Laboratory for Urban Sensing, Monitoring and Early Warning, Guangzhou 510060, China

Corresponding author: Xingdong Deng (dengxd_updsrig@163.com)

ABSTRACT Point clouds derived from LiDAR (Light Detection and Ranging) and photogrammetry systems are used to extract building footprints in dense urban areas. Two extraction methods based on DSM (Digital Surface Model) images and point clouds are comprehensively evaluated and compared. Firstly, photogrammetric point clouds are generated from aerial images of downtown Guangzhou, China, and compared with corresponding LiDAR point clouds. Then, DSM images are created using these point clouds and a threshold segmentation method is applied for building extraction. Although regularized buildings can be extracted according to the selection of appropriate height thresholds for the LiDAR DSM and photogrammetric DSM, blurry building boundaries exist for results of photogrammetric DSM when high trees are available nearby. LiDAR DSM extraction performs better in terms of Precision, Recall, and F-score metrics. A DoN (Difference of Normals) approach based on point cloud datasets is also quantitatively and qualitatively demonstrated. Our experiments show that when a suitable radius threshold is selected, the method provides satisfactorily normal calculation results and can successfully isolate building roofs from other objects in densely built-up areas. The majority of building extraction results have a precision > 0.9 and favorable Recall and F-score results. There is high consistency between photogrammetric and LiDAR point clouds. Although LiDAR provides higher extraction accuracy, photogrammetry is also useful for its more convenient acquisition and higher point cloud densities.

INDEX TERMS Photogrammetry, LiDAR, building extraction, digital surface model, difference of normals.

I. INTRODUCTION

The identification and extraction of buildings have become crucial issues in many applications, such as urban basic geodatabase updating, city planning management, disaster assessment, digital mapping, transportation planning, cadastral management, acoustic and energy studies, and telecommunication network design [1]–[4]. Collecting building information by field survey is labor-intensive and time-consuming. Building information updates occur slowly compared to the rapid rate of urbanization, especially in developing countries. To accommodate the demands of

The associate editor coordinating the review of this manuscript and approving it for publication was John Xun Yang[ID].

various applications, rapid, economical, and accurate building extraction is required. Nowadays, building extraction from remote sensing data has received research attention as it is rapid and cost-effective, and effective at large scales [5]–[8]. For a long time, automatic approaches to building detection have been difficult if not impossible due to scene complexity, incomplete extraction, and sensor dependencies, especially in big cities with dense buildings [9]–[11].

Methodologically, *building extraction* refers to the task of dividing a given dataset into non-overlapping homogeneous regions and recognizing the buildings from those regions. Various image-recognition algorithms have been proposed based on pixel features, geometric structure features, and object-based identification. Pixel-based classification

methods mainly analyze the features of each pixel in different spectral channels, extract a large number of features from the pixels, and then classify them. The level of recognition accuracy is mainly determined by the features extracted from the pixels and the classification method used, which include the parallelepiped, minimum distance, maximum likelihood, neural network, support vector machine, K-means and ISO-DATA clustering methods [12]. Besides pixel-based methods, geometric structure fitting has also been applied to extraction due to the regularity of buildings. The edges, corners, and contours of the buildings in an image can be extracted and a feature model matched and identified. Based on the geometric features, Brédif *et al.* provided a fully automatic global optimization framework to extract polygonal building footprints from DSM (Digital Surface Model) [13]. Their experiments proved that the contour vectorization accuracy was high. Recently, object-based building identification methods have emerged. In this process, an image is first divided into multiple object patches, which include texture, shape, and spectral features, using multi-scale segmentation. Then, a reliable method is selected to complete the classification process. Baatz *et al.* proposed a method based on a combination of spectral, texture, and context features and obtained good results with high-resolution remote sensing images [14]. The main aspect of object-oriented classification and recognition is to fully integrate the building's spectral, geometric, texture, and context features. Nowadays, deep learning techniques are widely used because they can automatically utilize large amounts of features to obtain high accuracy. Lin *et al.* proposed an efficient network used in segmentation of remote sensing images and achieved competitive results with much lower number of parameters and faster inference speed [15]. Chen *et al.* proposed a dense residual neural network (DR-Net), using a deeplabv3+Net encoder/decoder backbone with densely connected convolution neural network (DCNN) and residual network (ResNet) structure [16]. Those related studies showed that deep learning could improve the extraction accuracy. However, for the method, there is a high requirement for the hardware and labor investment (training samples) and technique training. Moreover, the building edges obtained from images are often thick and noisy and require post-processing to obtain thin and sharp boundaries. Abdollahi *et al.* introduced an end-to-end convolutional neural network called Generative Adversarial Network (GAN) to extract accurate building boundary [17]. Other researches had also designed specific convolution features network to refine the building contour [18]–[21]. However, most of the studies were tested in theoretical environment with high spatial resolution nadir remotely sensed images. It is difficult to reach a satisfied effect in the dense building area. Mostly, a polygonization step that converts building pixels into polygons is used by imposing *a priori* building properties that are manually defined and automatically tuned [22]. This strategy is usually adapted to building extraction in the densely-built areas of big cities, because there are usually obstructions from surrounding

buildings and it is virtually unavoidable even in very high-spatial resolution remote sensing images.

With the development of 3D scanners and the availability of point cloud data, building extraction has been improved. This has motivated a move towards using point clouds to extract building features. There are two classical methods that are widely used. The first employs the region growing technique, while the other delineates regions by detecting the edges in a dataset. The region growing approach starts by selecting a seed point, calculating its properties, and comparing them with those of adjacent points based on a certain connectivity measure to form a region. One drawback of region growing is that it usually fails when transition between two regions is smooth and difficult to distinguish by threshold parameters. It only works well when the initial seed points are noise-free and it is prone to excessive growth [23]. Besides the region growing method, edge extraction has been studied by many researchers [24]. Dorninger and Pfeifer used mean shift segmentation to detect buildings and used 2D-shape generalization to extract initial roof outlines from a point cloud obtained by airborne LiDAR (Light Detection and Ranging) [25]. Sampath and Shan modified a convex hull algorithm to extract building boundaries from raw point cloud data and applied hierarchical least-squares analysis to regularize the building outlines [26]. However, edge-based methods are susceptible to outliers and incomplete edges that do not form explicit segments. Among model-based building extraction methods, a robust method is the RANdom SAmple Consensus (RANSAC) approach [27], [28]. It randomly and iteratively samples the least number of data points necessary to determine the model parameters. Several researchers have presented the RANSAC paradigm for roof plane segmentation [29]–[31]. However, RANSAC is prone to finding pseudo-planes and its computational efficiency decreases significantly as the amount of point cloud data increases [32]. Classification or clustering techniques can also be used for the segmentation of LiDAR points. A feature vector is defined to characterize the object to be extracted as uniquely as possible. References [33], [34] demonstrate the use of clustering techniques for building extraction. In cases with planar surfaces, the feature vector at each point can consist of the surface normal and the location of the point [35]. A normal vector can be generated by selecting a neighborhood around a selected location and fitting a plane based on the least-squares method. Ioannou *et al.* proposed defining a multi-scale operator for unorganized point clouds directly using the estimated surface normal map of an unorganized point cloud [36]. This works well for object extraction from LiDAR point clouds.

LiDAR can provide accurate 3D point clouds for building extraction; however, airborne LiDAR acquisitions remain very costly, especially in big cities with complex surroundings. Typical commercial aerial LiDAR acquisitions cost at least \$20,000 per flight regardless of study area size [37], representing a significant barrier to its widespread application [38]. Moreover, there are various roof types in urban

areas, making it difficult to achieve automatic building detection in complex scenes. Many existing algorithms are intricate and often fail in very complex inner-city environments without enough points [39]. To overcome the cost and logistical barriers to routine and frequent acquisition of high-spatial-resolution 3D datasets, aerial photogrammetric point clouds can be used. High-spatial-resolution 3D point clouds can be produced by applying the SfM (Structure from Motion) method to large areas [40], [41]. In particular, methods of dense point-cloud generation (dense image matching) are increasingly available for professional and amateur applications, such as 3D modeling and mapping, robotics, medical imaging, surveillance, tracking, and navigation [42]. Nevertheless, the reliability of photogrammetric point clouds for building extraction need be evaluated because of the existence of noisy points [43].

Several recent studies have compared LiDAR and photogrammetry techniques based on factors such as accuracy, resolution, and dense 3D reconstructions of small scenes [38], [44]. However, only a few have reported their differences when applied to the extraction of real, dense urban buildings. Here, we demonstrate and evaluate a practical method of urban building extraction in Guangzhou, China. With its rapid development, the city urgently requires building information for urban spatial planning, land use management, disaster prevention, and emergency management. Section 2 introduces the study area and 3D point cloud datasets obtained from LiDAR scanning and image-based matching methods. The methods and results are presented in Sections 3 and 4, respectively. In Section 5, the point clouds obtained from the aerial photogrammetric dense matching method are analyzed and evaluated in detail. Finally, some conclusions are presented regarding the application of photogrammetric dense matching point clouds to urban building extraction.

II. STUDY AREA AND DATA

A. STUDY AREA

The study area was Guangzhou (23° 6' N, 113° 45' E), southern China. Guangzhou is located at the north-central edge of the Pearl River Delta facing Hong Kong and Macau and is one of the most important transportation hubs in southern China. At the end of 2018, the permanent resident population was about 14 million according to demographic inventory. The city presented a high urban density and buildings with diverse and complex sizes and shapes. The average relative humidity was 77% and annual rainfall was about 1736 mm. The abundant rainfall and heat benefit the growth of plants, but clouds sometimes make aerial photography difficult. It is difficult to obtain consecutive clear days for aerial data collection. Additionally, the presence of high trees makes it hard to scan complete buildings, especially their facades.

The study area was located in the central business district (Figure 1). This area contains several kinds of commercial and residential buildings. LiDAR and aerial oblique

FIGURE 1. Aerial photograph of the three study areas.

photogrammetric image datasets were obtained. Due to the cost of LiDAR and oblique photogrammetry data acquisition, only a few data with same coverage were obtained in the experiments. Three plots labeled A, B, and C were selected in this area for building-extraction purposes.

B. DATA

A set of aerial images and a LiDAR dataset covering an area of approximately 10 km^2 were available. Data acquisition was carried out by a Bell 407 helicopter flying at an average altitude of 1500 m above ground level in December 2016. A Leica RCD30 photogrammetry system was used, which comprised five 80-megapixel full-frame professional aerial cameras; one vertical and four oblique. Parallel flight paths were set in an east-west orientation with 159 m intervals between neighboring paths. There were large overlaps between adjacent strips to ensure data capture of building facades and other vertical surfaces. Such arrangements, along with oblique cameras, increase the sampling density of the captured surfaces. This pattern was designed to allow each area to be photographed from multiple angles. Using GPS accessories, image capture locations were recorded as meta-data in JPEG-formatted images. The specifications of the imaging sensor and resulting images are summarized in Table 1.

III. METHOD

A workflow was designed to compare the two types of data used for building extraction (Figure 2). Firstly, 3D scenes

TABLE 1. Specifications of the imaging sensor and acquired images.

Item	Value
Imaging sensor	Leica RCD30
Flight altitude	1500 m above ground level
Frame resolution	10,336 × 6732 pixels
Ground sample distance	5.8 cm
Focal lengths	83 mm (vertical) / 156 mm (oblique)
Along-track overlap	88%
Across-track overlap	82.6%
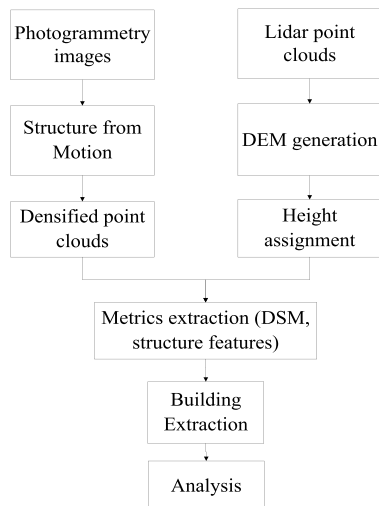
FIGURE 2. Workflow of the study.

FIGURE 3. (A) Sparse and (B) dense point clouds of the study area.

were reconstructed using photogrammetry images and SfM algorithms. Then, dense point clouds were generated using the dense matching method. These and LiDAR point clouds were preprocessed and features such as DSM and DoN (Difference of Normals) were extracted. Finally, the results were compared and evaluated.

A. AERIAL PHOTOGRAMMETRIC POINT CLOUDS

The photogrammetric point cloud data were generated from aerial images using the SfM algorithm. SfM computes an external camera pose for each image (indicating motion) and a 3D point cloud (indicating structure) to represent the pictured scene [45], [46]. The whole process yields a 3D point cloud (Figure 3a), as well as the camera poses, with re-projection residuals of 0.52 pixels. Then, the patch-match dense matching method was used to densify the point clouds. The patch-match dense matching method is an efficient patch-based stereo-matching plus depth-map refinement process that enforces consistency over multiple views [43]. For each image in the input image set, a reference image was selected to form a stereo pair for depth-map computation. Then, all of the depth maps were calculated. Since these raw depth maps generated by stereo vision may contain noise and

errors, each was refined by consistency checking using its neighboring depth maps. Finally, all the refined depth maps were merged to obtain the final dense matching point cloud (Figure 3b).

Computation was conducted using a desktop computer system (Table 2). The whole SfM process took 18 hours, including 1 hour for feature extraction, 1 hour for matching, and 16 hours for bundle adjustment. In this process, all of the images were calibrated and 1,510,309 sparse point clouds were generated. The dense matching took 3 hours and generated 127,228,537 points for the entire 10 km^2 area.

TABLE 2. Specifications of the computer used in the experiment.

Item	Value
OS	Ubuntu 14.04 (x86_64)
CPU	Intel® Xeon® E5-2603 v3 @ 1.60 GHz (12 cores)
RAM	125 GB
GPU	Nvidia Titan X

B. EXTRACTION OF METRICS

Points generated using photogrammetric techniques typically contained noise and errors. This complicates the estimation of metrics such as DTM (Digital Terrain Model), DSM, and so on, leading to erroneous values [47]. Here, a statistical method was used to trim noise that did not meet a certain criterion. Sparse outlier removal is based on computation of the distribution of point-to-neighbor distances in an input dataset [48]. By assuming that the resulting distribution is Gaussian with a mean of 50 and a standard deviation of 1, all points with mean distances that were outside an interval defined by the global distance mean and standard deviation were considered outliers and were trimmed from the dataset. After the noise-removal process, 90% of the points were retained. Although the point densities decreased after this process, they were still denser than those of the corresponding LiDAR point clouds (Table 3).

To obtain the metrics, the same procedures were carried out for the LiDAR and photogrammetric point clouds. Firstly, ground points and non-ground points were distinguished based on the cloth simulation filter (CSF) algorithm [49].

TABLE 3. Densities (spaces) of the LiDAR and photogrammetric point clouds (pts/m²).

Plot	LiDAR	Photogrammetric
A	4.74 (0.46)	37.5 (0.16)
B	6.42 (0.39)	36.61 (0.17)
C	9.76 (0.32)	29.28 (0.18)

Considering the point density and plot size, a cloth resolution of 10 m was chosen. The maximum iteration number was set to 500 and the classification threshold was 1. Height attribute was simply calculated as the difference between the elevation of the point and the terrain elevation estimated by the DTM. In this study, a DTM raster with a 0.5-m cell size was generated using ground points. Due to the irregular distribution of LiDAR data, the same size DSM raster was generated using non-ground point clouds based on the Delaunay triangulation (DT) algorithm. Quantitative statistics can be generated from point cloud and DSM data. Buildings can be extracted from DSM data based on the height distribution and prior knowledge. In our experiment, the height threshold method was applied for building extraction and certain metrics were selected for evaluation.

In addition to building extraction based on DSM images, a point cloud segmentation strategy called the DoN was also tested [36]. The concept of DoN defines a multi-scale operator directly using the estimated surface normal map of point clouds. The surface normals estimated at any given radius reflect the underlying geometry of the surface at the scale of the support radius. If the directions of the two surface normals are nearly identical, then the structure of the surface does not change significantly from the first radius to the second. If the structure of the larger neighborhood is significantly different from that of the smaller neighborhood, then the direction of the two estimated normal are likely to vary dramatically. In the extraction of building roofs, we can compare the response of the normal across two different radii: $r_1 < r_2$. In the process, the DoN is first calculated for each point within its multi-scale neighbors to separate the points based on the surface normal difference. Then, the DoNs of all points are clustered with the Euclidean distance threshold segmentation method. The final step of segmentation separates the planar and nonplanar segments based on their distances and connectivity, respectively. The calculation of the DoN operator $\Delta_{\hat{n}}$ for any point p in a point cloud P, is defined as:

$$\Delta_{\hat{n}}(p, r_1, r_2) = \frac{\hat{n}(p, r_1) - \hat{n}(p, r_2)}{2} \qquad (1)$$

where $r_1, r_2 \in R$, $r_1 < r_2$, and $\hat{n}(p, r)$ is the surface normal estimation at point p, given the support radius r. For a given r_1 and r_2, the result of applying the $\Delta_{\hat{n}}$ operator to all the points is a vector map, where a DoN vector is assigned to each point. Since each DoN is the normalized sum of two unit normal vectors, the magnitudes of the $\Delta_{\hat{n}}$ vectors are

always within [0,1]. In our building extraction, the DoN vectors were selected based on their magnitudes $\|\Delta_{\hat{n}}(p)\|$. After computation, a simple Euclidean distance threshold-based clustering algorithm [50] was applied with a distance tolerance to extract the buildings.

IV. RESULTS

A. EXTRACTION BASED ON DSM
The DSMs generated from the LiDAR and photogrammetric point clouds are shown in Figure 4. The photogrammetric DSM contains a clear saw-tooth effect (blurry building boundaries). The LiDAR DSM is clear and represents the building boundaries accurately. Although the structure patterns are similar, the LiDAR DSM has a clear and descriptive structure. This is because there were many noisy and outlying data points in the photogrammetric point clouds, even after the noise-removal process. Although there is a higher density in the photogrammetric point clouds, their geometric accuracy is lower than that of LiDAR. Spectral or color information is an advantage of photogrammetric point clouds that is not available in LiDAR data. However, this mostly only improves the visualization effect.
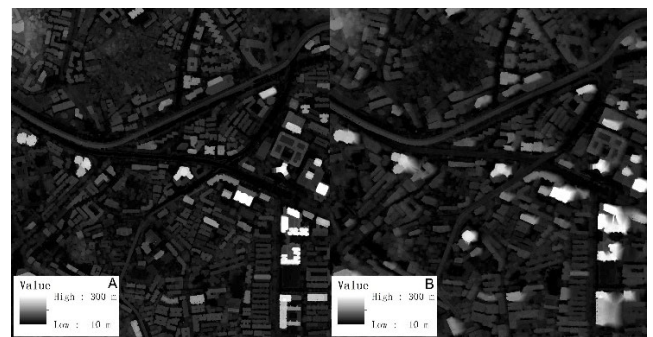
FIGURE 4. DSM obtained from (A) LiDAR and (B) photogrammetric point clouds.

DSM is a priority feature for extracting building boundaries, and a straightforward threshold method was applied. In order to evaluate the results, the building truth was labeled manually using the GIS processing software QGIS (v3.4). The Precision, Recall, and *F*-score performance metrics were calculated for each of the three selected sections and the extraction results are shown in Figures 5, 6, and 7.

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (3)$$

$$F_{Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (4)$$

where TP is the number of true positives; FP is the number of false positives; and FN is the number of false negatives.

Based on prior knowledge obtained from the urban survey, heights of the trees were between 5 m to 25 m, so height thresholds of 5 m to 25 m were selected for comparison.
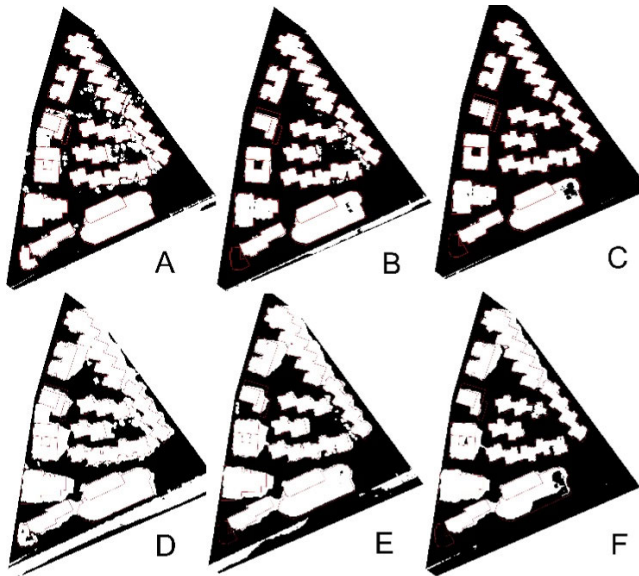
FIGURE 5. building extraction result of section A from LiDAR (A, B, C) and photogrammetric (D, E, F) DSMs based on height thresholds of 15 m (A, D), 20 m (B, E), and 25 m (C, F).

FIGURE 7. building extraction results of section C from LiDAR (A, B, C) and photogrammetric (D, E, F) DSMs based on height thresholds of 15 m (A, D), 20 m (B, E), and 25 m (C, F).

FIGURE 6. building extraction results of section B from LiDAR (A, B, C) and photogrammetric (D, E, F) DSMs based on height thresholds of 15 m (A, D), 20 m (B, E), and 25 m (C, F).
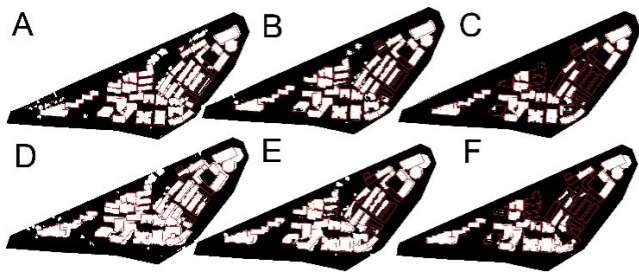
Regularized buildings could be extracted when a 25-m threshold was selected in section A (Figures 5), B (Figures 6) and section C (Figures 7). However, the building boundaries of the photogrammetric DSM were blurry, especially when high trees or other buildings were nearby. Indeed, the selected threshold may greatly affect the extraction result.

When a low height threshold was chosen, true positive numbers improved greatly and the Precision score of extraction was also high. However, the Recall and F-scores were low when a low threshold was chosen. With the 5-m threshold, nearly all of the buildings could be extracted. However, trees and other objects were also classified as buildings. Misclassification was lesser with the 25-m threshold, although a few buildings could not be extracted. So, it is difficult to select the optimal threshold for building extraction with this method without prior knowledge about the study area.

The effects of the threshold setting on precision of section A are shown in Figure 8. It shows that the F-score peaks at a threshold of 25 m and then decreases, while Recall peaks at 25 m and remains high at greater thresholds. It also shows that
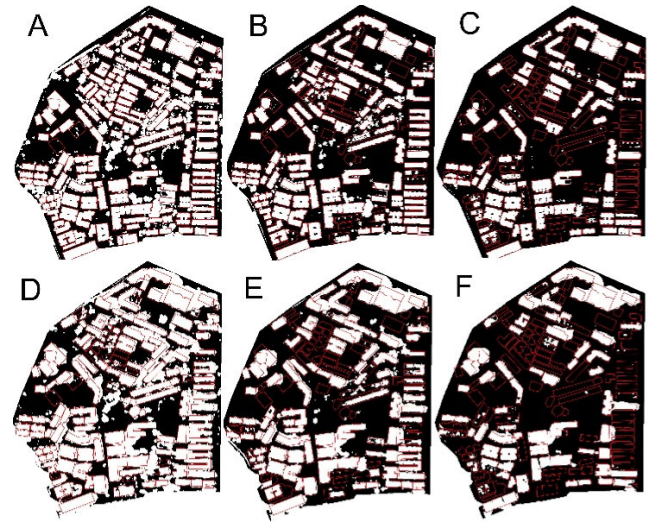
the F-score of photogrammetric DSM extraction is always lower than that of LiDAR DSM extraction regardless of the threshold selected.

B. EXTRACTION BASED ON POINT CLOUDS

In the DoN implementation, the small radius (r_1) and large radius (r_2) were set to 1 m and 10 m, respectively. Such DoN parameters settings have been found to provide good isolation of points in urban LiDAR scenes. Applying the Euclidean cluster extraction method to the resulting point cloud, building roofs were clearly clustered with the scene. For segmentation, a threshold value of 0.1 was applied for building roofs and 0.4 for trees. The building roof extraction results are shown in Figures 9 and 10. It shows that most of the buildings could be successfully extracted from both types of point clouds. After comparison with the true building boundaries of sections A, B, and C, we find that the FP ratio is rather low.

An evaluation was also carried out to compare DoN segmentation with the two types of data. Table 4 illustrates the results of our evaluation in the form of a Precision/Recall/ F-score over ground truth objects. For each cluster, the point classification was compared with each of the ground truth labels. It was found that the majority of the results had a precision > 0.9. The Recall and F-score results also appear favorable. Overall, the LiDAR extraction results have some advantages over the photogrammetric ones in terms of F-scores.

V. DISCUSSION

In this study, photogrammetric point clouds were obtained indirectly from images. This required a process of both aerial triangulation and dense matching. Many factors, such as feature selection, corresponding matches, and patch-matching
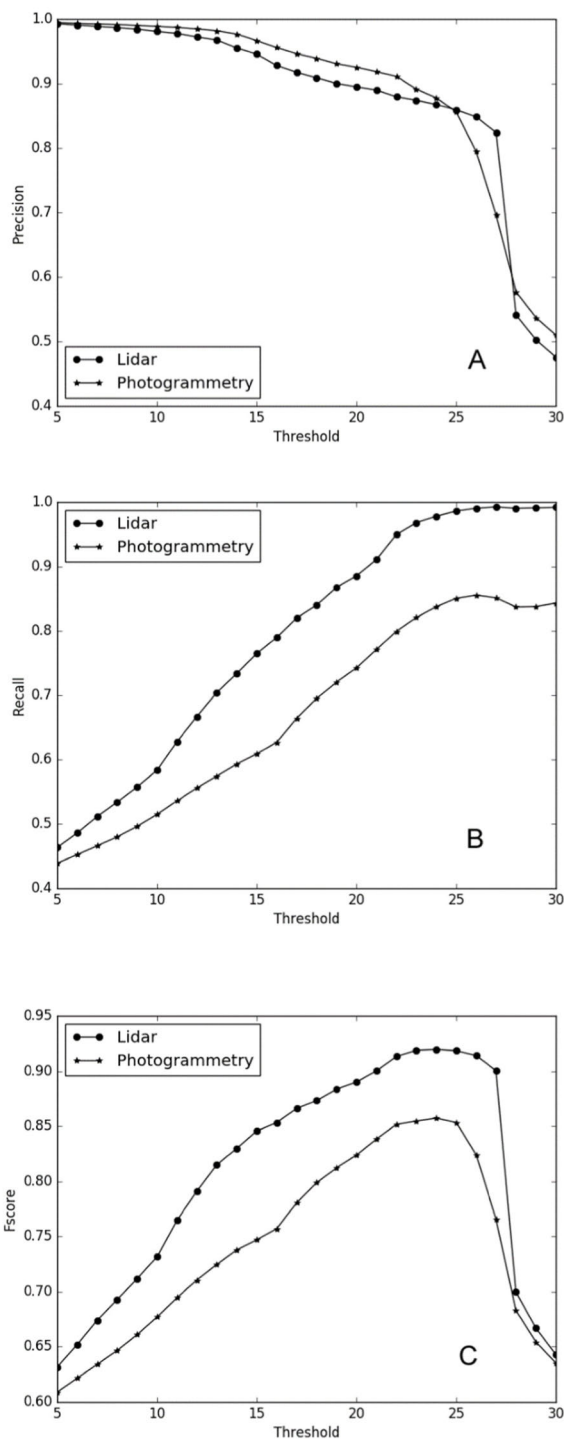
FIGURE 8. Precision (A), Recall (B) and *F*-scores (C) in relation to height threshold for LiDAR and photogrammetric DSM extraction results.

FIGURE 9. Building extraction based on LiDAR point clouds for sections A, B, and C.

affect the quality of the point clouds. However, LiDAR can directly measure distances so, theoretically, fewer factors affect the geometric accuracy. Although the point density of photogrammetric point clouds was high, there was low 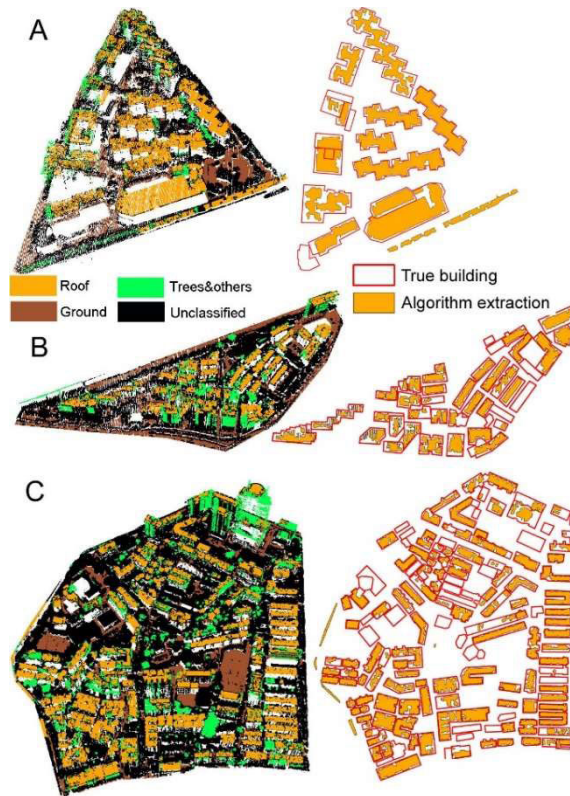geometric accuracy due to greater noise. Compared to the photogrammetric point clouds, the LiDAR point clouds had better accuracy and structure.

In this study, DSMs were calculated to extract buildings. A closer look at the DSMs revealed that photogrammetric DSMs contained noise and outliers along building boundaries. These are the main causes of the saw-tooth appearance of the photogrammetric DSM. Such artifacts degrade the quality of photogrammetric DSMs and hamper their reliable usage in urban building extraction. Here, buildings were extracted based on DSM thresholds for both LiDAR- and photogrammetry-generated point clouds. Although their structural patterns were similar, the LiDAR DSM had a clearer and more descriptive structure. Compared with photogrammetric DSM, LiDAR DSM appears to provide a better building extraction result. After evaluation, it was found that regularized buildings could be extracted with different thresholds selected for different sections. Moreover, building boundaries were blurry, especially when high trees, nearby buildings, and other relief displacements existed. When a low height threshold was chosen, numbers of true positives improved greatly and extraction precision was also high. However, the Recall and *F*-scores were low when a low threshold was chosen. So, it is difficult to select the optimal DSM threshold for building extraction. It was also found that the *F*-score of photogrammetric DSM extraction was always lower than that of LiDAR DSM due to various influences.
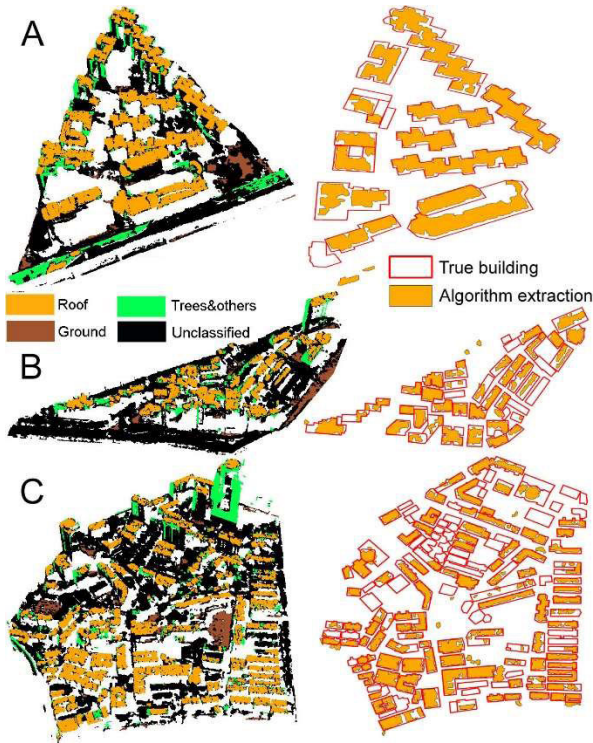
FIGURE 10. Building extraction based on photogrammetric point clouds for sections A, B, and C.

TABLE 4. Classification results and performance metrics.

LiDAR	A	55,057	6408	17,197	0.90	0.76	0.82
	B	70,322	1459	57,991	0.98	0.55	0.70
	C	409,253	15448	393,732	0.96	0.51	0.67
Photogramm etry	A	471,750	7273	235,067	0.98	0.67	0.80
	B	482,225	18717	500,880	0.96	0.49	0.65
	C	1,423,395	70534	1,440,046	0.95	0.50	0.65

Building roof extraction based on DSM could not comprehensively utilize the 3D information, so the algorithm based on point clouds was also tested for both LiDAR and photogrammetric data. In this study, the DoN segmentation strategy was shown to be surprisingly powerful in extracting point features according to their scale. Selecting the parameters r_1 and r_2 for DoN may cause a large response in the surface of interest. Selection of the neighborhood affects the calculation of the normals and, hence, the segmentation results. In the extraction of building roofs, we compared the responses of the normals across two different radii $r_1 < r_2$. A low threshold may result in a planar region being classified as trees, while a high one will cause the opposite. Our experiments with different buildings found that using $r_1 = 1$ m and $r_2 = 10$ m gives satisfactory results [36]. These small radii were chosen because there were enough neighbor points to calculate point normals. This can provide a good estimation of the surface normal. It can balance the two types of errors—false positives and false negatives. Such DoN parameter settings were found to provide good isolation of points in urban LiDAR scenes.

Appropriate parameter selection can maximize the difference in the DoN magnitudes of building roofs and other objects.

After applying the euclidean cluster extraction method to the resulting point cloud, clear clustering could be generated in a scene. For each point clouds cluster, a threshold value of 0.1 m was applied for building roof planar fitting. A low measure threshold (0.1) yielded horizontal and planar surfaces that were mostly classified as buildings. On the other hand, a high value (0.4) yielded rough or vertical surfaces, which may indicate that the points represent trees or the vertical facets of buildings. The potential-based clustering approach can reduce the effect of such off-center outliers that are mixed in with the data points, and correctly determine the numbers and locations of these clusters. The cluster extraction algorithms do not necessarily take into account the diversity of the geospatial nature of LiDAR datasets, such as returns from trees, vertical walls, chimneys, curved structures, and other irregular objects, in addition to the noise present in the data. It should be noted that curved building surfaces can lead to over-segmentation. In this scenario, the LiDAR points of a curved surface will be segmented into more than one planar piece within a certain tolerance.

The segmentation quality was quantitatively evaluated on the photogrammetric and LiDAR point cloud datasets. Building roofs were automatically segmented from these datasets. It was shown that the majority of the results had a precision > 0.9, and the Recall and F-score results appear favorable. Overall, the LiDAR extraction results have some advantages over the photogrammetric ones, considering the comprehensive F-score metric.

VI. CONCLUSION

In this paper, photogrammetric and LiDAR point clouds were used to extract building data from aerial imagery. Two methods based on DSM images and point clouds were tested. The comprehensive analysis showed good consistency between the two types of data. However, compared to LiDAR data, photogrammetric point clouds provide poor building extraction accuracy. There were some differences between the LiDAR and photogrammetric methods. From a practical point of view, the trade-off between effective cost and extraction accuracy should be exhaustively considered. Compared with DSM image extraction, using DoN as a multi-scale operator can obtain the advantages of 3D point clouds. Applying the methods to LiDAR and photogrammetric data of real urban areas qualitatively demonstrated the effectiveness of DoN segmentation in classifying building roofs.

It should be noted that photogrammetric point clouds provide lower geometric accuracy than LiDAR ones. However, the point density of photogrammetric point clouds is much higher and may include much more redundant data. The DoN was proven to be a theoretically sound and practically effective technique for satisfactorily detecting these nonplanar points. However, the size of the neighborhood must be specified for surface normal estimation.

In summary, when extracting buildings from imaging datasets, photogrammetric point clouds are a good option if LiDAR data are unavailable. However, for urban building extraction based on photogrammetric point clouds, multi-view images should be considered, and noise removal should be carried out. At the same time, color information could be used in future to improve the accuracy of extracting the related metrics. Future work should exploit the DoN scale operator over several radii for building extraction and integrate it with cluster-recognition methods.

REFERENCES

[1] R. Huang, B. Yang, F. Liang, W. Dai, J. Li, M. Tian, and W. Xu, "A top-down strategy for buildings extraction from complex urban scenes using airborne LiDAR point clouds," *Infr. Phys. Technol.*, vol. 92, pp. 203–218, Aug. 2018.

[2] B. Benciolini, V. Ruggiero, A. Vitti, and M. Zanetti, "Roof planes detection via a second-order variational model," *ISPRS J. Photogramm. Remote Sens.*, vol. 138, pp. 101–120, Apr. 2018.

[3] S. Du, Y. Zhang, Z. Zou, S. Xu, X. He, and S. Chen, "Automatic building extraction from LiDAR data fusion of point and grid-based features," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 294–307, Aug. 2017.

[4] R. Cao, Y. Zhang, X. Liu, and Z. Zhao, "Roof plane extraction from airborne LiDAR point clouds," *Int. J. Remote Sens.*, vol. 38, no. 12, pp. 3684–3703, Jun. 2017.

[5] X. Lai, J. Yang, Y. Li, and M. Wang, "A building extraction approach based on the fusion of LiDAR point cloud and elevation map texture features," *Remote Sens.*, vol. 11, no. 14, p. 1636, Jul. 2019.

[6] E. Næsset, "Practical large-scale forest stand inventory using a small-footprint airborne scanning laser," *Scand. J. Forest Res.*, vol. 19, no. 2, pp. 164–179, 2004.

[7] S. E. Reutebuch, H.-E. Andersen, and R. J. McGaughey, "Light detection and ranging (LiDAR): An emerging tool for multiple resource inventory," *J. Forestry*, vol. 103, no. 6, pp. 286–292, 2005.

[8] R. W. Kulawardhana, S. C. Popescu, and R. A. Feagin, "Fusion of LiDAR and multispectral data to quantify salt marsh carbon stocks," *Remote Sens. Environ.*, vol. 154, pp. 345–357, Nov. 2014.

[9] G. Sohn and I. Dowman, "Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 1, pp. 43–63, 2007.

[10] F. Rottensteiner and C. Briese, "A new method for building extraction in urban areas from high-resolution LiDAR data," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 34, pp. 295–301, Sep. 2001.

[11] S. L. Ullo, C. Zarro, K. Wojtowicz, G. Meoli, and M. Focareta, "LiDAR-based system and optical VHR data for building detection and mapping," *Sensors*, vol. 20, no. 5, p. 1285, Feb. 2020.

[12] S. Raschka, *Python Machine Learning*. Birmingham, U.K.: Packt, 2015.

[13] M. Brédif, O. Tournaire, B. Vallet, and N. Champion, "Extracting polygonal building footprints from digital surface models: A fully-automatic global optimization framework," *ISPRS J. Photogramm. Remote Sens.*, vol. 77, no. 1, pp. 57–65, 2013.

[14] M. Baatz and A. Schape, "Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation," *Angew. Geographische Inf.-Verarbeitung*, vol. 12, pp. 12–23, Jan. 2000.

[15] J. Lin, W. Jing, H. Song, and G. Chen, "ESFNet: Efficient network for building extraction from high-resolution aerial images," *IEEE Access*, vol. 7, pp. 54285–54294, 2019.

[16] Y. Liu, J. Zhou, W. Qi, X. Li, L. Gross, Q. Shao, Z. Zhao, L. Ni, X. Fan, and Z. Li, "ARC-Net: An efficient network for building extraction from high-resolution aerial images," *IEEE Access*, vol. 8, pp. 154997–155010, 2020.

[17] A. Abdollahi, B. Pradhan, S. Gite, and A. Alamri, "Building footprint extraction from high resolution aerial images using generative adversarial network (GAN) architecture," *IEEE Access*, vol. 8, pp. 209517–209527, 2020.

[18] J. Yang, B. Price, S. Cohen, H. Lee, and M.-H. Yang, "Object contour detection with a fully convolutional encoder-decoder network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 193–202.

[19] Y. Wang, X. Zhao, Y. Li, and K. Huang, "Deep crisp boundaries: From boundaries to higher-level tasks," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1285–1298, Mar. 2019.

[20] A. Kelm, V. S. Rao, and U. Zölzer, "Object contour and edge detection with RefineContourNet," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, 2019, pp. 246–258.

[21] T. Lu, D. Ming, X. Lin, Z. Hong, X. Bai, and J. Fang, "Detecting building edges from high spatial resolution remote sensing imagery using richer convolution features network," *Remote Sens.*, vol. 10, no. 9, p. 1496, Sep. 2018.

[22] R. Deng, C. Shen, S. Liu, H. Wang, and X. Liu, "Learning to predict crisp boundaries," 2018, *arXiv:1807.10097*. [Online]. Available: http://arxiv.org/abs/1807.10097

[23] A.-V. Vo, L. Truong-Hong, D. F. Laefer, and M. Bertolotto, "Octree-based region growing for point cloud segmentation," *ISPRS J. Photogramm. Remote Sens.*, vol. 104, pp. 88–100, Jun. 2015.

[24] S. R. Lach and J. P. Kerekes, "Robust extraction of exterior building boundaries from topographic LiDAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 2, Jul. 2008. p. II-85.

[25] P. Dorninger and N. Pfeifer, "A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds," *Sensors*, vol. 8, no. 11, pp. 7323–7343, Nov. 2008.

[26] 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.

[27] H. Li, J. Qin, X. Xiang, L. Pan, W. Ma, and N. N. Xiong, "An efficient image matching algorithm based on adaptive threshold and RANSAC," *IEEE Access*, vol. 6, pp. 66963–66971, Nov. 2018.

[28] C. Shao, C. Zhang, Z. Fang, and G. Yang, "A deep learning-based semantic filter for RANSAC-based fundamental matrix calculation and the ORB-SLAM system," *IEEE Access*, vol. 8, pp. 3212–3223, 2020.

[29] D. Chen, L. Zhang, P. T. Mathiopoulos, and X. Huang, "A methodology for automated segmentation and reconstruction of urban 3-D buildings from ALS point clouds," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 10, pp. 4199–4217, Oct. 2014.

[30] D. Chen, L. Zhang, J. Li, and R. Liu, "Urban building roof segmentation from airborne LiDAR point clouds," *Int. J. Remote Sens.*, vol. 33, no. 20, pp. 6497–6515, May 2012.

[31] Y. Xu, S. Tuttas, L. Hoegner, and U. Stilla, "Geometric primitive extraction from point clouds of construction sites using VGS," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 424–428, Mar. 2017.

[32] R. Schnabel, R. Wahl, and R. Klein, "Efficient ransac for point-cloud shape detection," *Comput. Graph. Forum*, vol. 26, no. 2, pp. 214–226, 2007.

[33] S. Filin and N. Pfeifer, "Segmentation of airborne laser scanning data using a slope adaptive neighborhood," *ISPRS J. Photogramm. Remote Sens.*, vol. 60, no. 2, pp. 71–80, 2006.

[34] J. Shan and A. Sampath, "Building extraction from LiDAR point clouds based on clustering techniques," in *Topographic Laser Ranging and Scanning*. Boca Raton, FL, USA: CRC Press, Nov. 2008, ch. 15, pp. 421–444.

[35] R. Hoffman and A. K. Jain, "Segmentation and classification of range images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 608–620, Sep. 1987.

[36] Y. Ioannou, B. Taati, R. Harrap, and M. Greenspan, "Difference of normals as a multi-scale operator in unorganized point clouds," in *Proc. 2nd Int. Conf. 3D Imag., Modeling, Process., Vis. Transmiss.*, Oct. 2012, pp. 501–508.

[37] T. L. Erdody and L. M. Moskal, "Fusion of LiDAR and imagery for estimating forest canopy fuels," *Remote Sens. Environ.*, vol. 114, no. 4, pp. 725–737, Apr. 2010.

[38] J. P. Dandois and E. C. Ellis, "High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision," *Remote Sens. Environ.*, vol. 136, pp. 259–276, Sep. 2013.

[39] F. Rottensteiner, G. Sohn, M. Gerke, J. D. Wegner, U. Breitkopf, and J. Jung, "Results of the ISPRS benchmark on urban object detection and 3D building reconstruction," *ISPRS J. Photogramm. Remote Sens.*, vol. 93, pp. 256–271, Jul. 2014.

[40] H. Aliakbarpour, V. B. S. Prasath, K. Palaniappan, G. Seetharaman, and J. Dias, "Heterogeneous multi-view information fusion: Review of 3-D reconstruction methods and a new registration with uncertainty modeling," *IEEE Access*, vol. 4, pp. 8264–8285, 2016.

[41] A. G. Melo, M. F. Pinto, L. M. Honorio, F. M. Dias, and J. E. N. Masson, "3D correspondence and point projection method for structures deformation analysis," *IEEE Access*, vol. 8, pp. 177823–177836, 2020.

[42] F. Remondino, M. G. Spera, E. Nocerino, F. Menna, and F. Nex, "State of the art in high density image matching," *Photogramm. Rec.*, vol. 29, pp. 144–166, Jun. 2014.

[43] S. Shen, "Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1901–1914, May 2013.

[44] S. Alexiou, G. Deligiannakis, A. Pallikarakis, I. Papanikolaou, E. Psomiadis, and K. Reicherter, "Comparing high accuracy t-LiDAR and UAV-SfM derived point clouds for geomorphological change detection," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 6, p. 367, May 2021.

[45] P. Moulon, P. Monasse, and R. Marlet, "Adaptive structure from motion with a contrario model estimation," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 257–270.

[46] P. Moulon, P. Monasse, and R. Marlet, "Global fusion of relative motions for robust, accurate and scalable structure from motion," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3248–3255.

[47] K. Zhang, S.-C. Chen, D. Whitman, M.-L. 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.

[48] R. Rusu, "Fast 3D recognition and pose using the viewpoint feature histogram," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 2155–2162.

[49] W. Zhang, J. Qi, P. Wan, H. Wang, D. Xie, X. Wang, and G. Yan, "An easy-to-use airborne LiDAR data filtering method based on cloth simulation," *Remote Sens.*, vol. 8, no. 6, p. 501, Jun. 2016.

[50] R. B. Rusu, "Semantic 3D object maps for everyday manipulation in human living environments," *KI-Künstliche Intelligenz*, vol. 24, no. 4, pp. 345–348, 2010.

LIANG GUO received the Ph.D. degree in cartography and geographic information systems from the University of Chinese Academy of Sciences, Beijing, China, in 2011. He is currently the Minister of Research and Development Fellow with Guangzhou Urban Planning and Design Survey Research Institute. His research interests include spatial data mining and airborne SLAM.

XINGDONG DENG received the Ph.D. degree in engineering with South China University of Technology, Guangzhou, China, in 2011. He is currently the President of Guangzhou Urban Planning and Design Survey Research Institute and an Adjunct Professor of South China University of Technology. He received the Special Government Allowances Expert of the State Council and was awarded "Guangzhou Excellent Expert," in 2012. He was included in the second echelon of 121 Talent Eechelon Project, in 2014. His research interests include urban planning, urban transportation, urban design, geospatial technology, and geotechnical engineering.

YANG LIU received the Ph.D. degree in cartography and geographic information system from Wuhan University, in 2008. He is currently the Head of the Technology Department, Guangzhou Urban Planning and Design Survey Research Institute. He also have the title of registered surveyor, registered planner, and professional engineer. His research interests include spatial data mining and disaster prediction.

HUAGUI HE received the B.S. degree in cartography from Wuhan University, in 2000. He is currently the Director of the Geographic Information Center, Guangzhou Urban Planning and Design Survey Research Institute. He is also a professional senior engineer in surveying and mapping engineering. His research interests include smart city and computer cartography.

HONG LIN is currently the Vice President of Guangzhou Urban Planning and Design Survey Research Institute. He is a professor in engineering investigation, the first batch of national registered surveyors engineering and the "special government allowances expert" in China. He has been worked on surveying for 34 years. His research interests include engineering survey and design. As the project leader, he won many significant science and technology awards, including three national excellent engineering survey and design industry awards.

GUANGXIN QIU graduated from Tongji University, in 1988. He has been a Professional Senior Engineer with Guangzhou Urban Planning and Design Survey Research Institute, since 2000. He completed Guangzhou underground pipeline survey project. He was the major participant of the "travelling measurement system of urban street view" project by using vehicle-mounted multi-sensor. He was the main editor of national and industrial standards, such as the urban underground pipeline detection. He won three provincial technology progress awards.

WEIJUN YANG received the Ph.D. degree in cartography and geographic information systems from the University of Chinese Academy of Sciences, Beijing, China, in 2008. He is currently the Chief Engineer of Guangzhou Urban Planning and Design Survey Research Institute. He received the professional title of senior engineer, in 2018. His research interests include geographic information system development and spatial data processing.

● ● ●