Research Article

Qiuji Chen, Xin Wang*, Mengru Hang, and Jiye Li

Research on the improvement of single tree segmentation algorithm based on airborne LiDAR point cloud

https://doi.org/10.1515/geo-2020-0266
received January 19, 2021; accepted May 19, 2021

Abstract: The correct individual tree segmentation of the forest is necessary for extracting the additional information of trees, such as tree height, crown width, and other tree parameters. With the development of LiDAR technology, the research method of individual tree segmentation based on point cloud data has become a focus of the research community. In this work, the research area is located in an underground coal mine in Shenmu City, Shaanxi Province, China. Vegetation information with and without leaves in this coal mining area are obtained with airborne LiDAR to conduct the research. In this study, we propose hybrid clustering technique by combining DBSCAN and K-means for segmenting individual trees based on airborne LiDAR point cloud data. First, the point cloud data are processed for denoising and filtering. Then, the pre-processed data are projected to the XOY plane for DBSCAN clustering. The number and coordinates of clustering centers are obtained, which are used as an input for K-means clustering algorithm. Finally, the results of individual tree segmentation of the forest in the mining area are obtained. The simulation results and analysis show that the new method proposed in this paper outperforms other methods in forest segmentation in mining area. This provides effective technical support and data reference for the study of forest in mining areas.

Keywords: single tree segmentation, LiDAR technology, mining area forest, DBSCAN, K-means clustering algorithm

*** Corresponding author: Xin Wang,** Department of Geography, College of Geomatics, Xi'an University of Science and Technology, Xi'an, 710054, China, e-mail: xinwangxkd@163.com
Qiuji Chen, Jiye Li: Department of Geography, College of Geomatics, Xi'an University of Science and Technology, Xi'an, 710054, China
Mengru Hang: Xi'an Institute of Geotechnical Investigation and Surveying Mapping, Xi'an, 710054, China

1 Introduction

The large-scale mining causes huge subsidence land around the mines, causing different degrees of damage to the surrounding natural landscape and ecological environment [1]. With the acceleration of China's ecological civilization, green mining has turned to be the development trajectory of the national mining industry [2]. Green mining is modern practice to mine in a scientific and environmentally friendly manner to have the least impact on the environment [3]. Green mine construction allows the overall environment of the mining area to be reasonably laid out to protect nature [2,3]. Green mining recognizes the importance of forests and therefore pursues a progressive path to eliminate or reduce undesirable impacts on forest ecosystems and biodiversity [4]. Therefore, we need to systematically study the impact of mining on the ecological environment, especially to evaluate the current situation of vegetation in mining areas. These can provide assistance and suggestions for the environmental assessment and management of mining areas [5].

Vegetation is the producer in an ecosystem and is a major player in the ecological environment. Any significant damage to the vegetation in the mining area reduces the stability and functionality of the ecosystem [6]. The vegetation in the mining area comprises forests with tree as a basic unit. The parameters of the single tree effectively showcase the impact of mining on vegetation [7]. Currently, individual tree segmentation mainly relies on the manual measurement and remote sensing technologies. It is noteworthy that the manual measurement methods take a long time for collecting data. Thus, it is difficult to collect the data at a large scale. In addition, these techniques also have a high monitoring cost [8]. Although, it is possible to efficiently obtain the data using traditional remote sensing methods, such as satellite imaging and photogrammetry, the satellite images have limits in terms of resolution and the data accuracy is low.

Photogrammetry only provides two-dimensional (2D) data, and its ability to obtain three-dimensional (3D) information of the forest is poor [9].

In recent years, the development of LiDAR technology created an opportunity to collect accurate data of single-tree segmentation conveniently. LiDAR incident busts of laser pulses on the vegetation and receives the echo signals that contain the desired information, such as 3D positional information of trees. The received reflected signals are then used to generate point clouds [10]. The point clouds have a wide range, high precision, and can be processed efficiently [11].

At present, many researchers have proposed numerous methods for segmenting individual trees using LiDAR data. For example, Hyyppa et al. started with local maximum and used a region growing algorithm to segment individual trees [12]. Wang et al. proposed a marker-controlled watershed segmentation algorithm that assumed that treetops were represented by local radiation maxima and used the crown points as markers to improve accuracy [13]. The aforementioned method is more suitable for the forest with a simple structure, but it is not suitable for taller trees and thicker forests. This method is difficult to correctly segment the cases where adjacent branches are intertwined [14]. Therefore, a new method for segmenting individual trees using an adaptive TWS (treetop window size) to find tree vertices was proposed, which can reduce the phenomenon of over-segmentation in broad-leaved forest [15,16]. For mixed forests, Koch et al. proposed that tree tops were detected with a local maximum filter, and then using a pouring algorithm to segment individual trees [14].

What these segmentation algorithms share is that they segment individual trees using canopy height model (CHM), which is a raster image interpolated from LiDAR points. These kinds of algorithms are not ideal enough, as the CHM can have inherent errors and uncertainties in the process of CHM generation, which will lead to decline in data quality and affect the accuracy of segmenting individual trees. For example, there may be calculation error in the process of original data filtering, and spatial error can be introduced during the interpolation process [17]. Therefore, it is extremely important to develop a new method to segment individual trees directly from the LiDAR point cloud. Reitberger et al. proposed a method using the Random Sample Consensus (RANSAC) for tree segmentation based on normalized cut segmentation in graph theory. This algorithm had high segmentation accuracy, but the normalized cut segmentation required too much computation, resulting in low segmentation efficiency [18]. On this basis, Ayrey et al. improved the normalized cut segmentation algorithm and proposed a new algorithm combining Layer stacking algorithm with DBSCAN, which greatly improved the efficiency and accuracy of segmenting individual trees [19]. In addition, some classical algorithms in the field of computer vision and image processing (K-means, region growing algorithm) have also been applied to segment individual trees based on LiDAR point cloud data [20,21].

In this study, we developed a new algorithm to segment individual trees based on the synergy of DBSCAN and K-means, which can directly segment LiDAR point cloud data. This new algorithm can well solve the problems that DBSCAN algorithm has poor processing ability for multi-dimensional data and K-means has high requirements for the initial clustering center. The algorithm first deletes the Z value of each point in the original point cloud data and uses the 2D DBSCAN algorithm for coarse clustering. The points after clustering are processed and their Z values are added. Then, we calculate the number of clusters and the average of the coordinates of all the points in each cluster, and take it as the initial value. Finally, we use K-means clustering algorithm to complete accurate single tree segmentation. Based on the LiDAR point cloud data, we perform segmenting individual trees for leafy and leafless forest located in the target area. We also verify the accuracy of the new algorithm. The results and analysis show that the new method improves the accuracy of segmenting individual trees. In addition, the proposed method also provides feasible technical support and effective data reference for the restoration of the ecological environment in the mining area and the construction of green mines.

2 Data acquisition and processing

2.1 Overview and explanation of data

The target area in this work is located near a coal mine in the northwestern part of Shenmu City, Shaanxi Province, China, on the edge of the Mu Us Sandy Land. The coal mine was officially put into production in December 2013. The mining area is 119.7735 km^2, the mining depth is 80–240 m, and the mining method is shaft mining. The main geomorphological types are loess hilly and gully areas and river valley terraces. The topography of the area is northwest, southwest, and central low. The target area is a typical mid-temperate, semi-arid, and continental monsoon climate, with an annual average temperature of 8.8°C and an average annual rainfall of 436.6 mm. The target area is dominated by psammophytic vegetation. The most abundant tree in the target area is salix mongolica. In recent years, large-scale mining in this area has caused damage to local primary vegetation.

2.2 Data collection

2.2.1 Acquisition of airborne LiDAR point cloud data

In this work, the point cloud data used are obtained by DJM600 UAV equipped with RIEGLminiVUX-1UVA laser scanning system. The measurement accuracy of this equipment is 15 mm. We collect LiDAR point cloud data from the target area in two phases, i.e., on October 22, 2019 and November 23, 2019. On these days, the weather was sunny. The tree leaves were existent on October 22, 2019, and the collected data were recorded as leafy phase data. However, on November 23, 2019, the tree leaves fell off, and the collected data were recorded as leafless phase data. The airborne LiDAR parameters and UAV flight parameters are presented in Table 1.

2.2.2 Field survey data collection

The ground survey and drone flight operations are conducted at the same time. We used drones to obtain high-resolution orthophoto images of the area. The ground truth is obtained by visual interpretation of the image. We randomly select five sample plots from the data of leafy stage, which includes 47 trees. Similarly, we also randomly select five sample plots from leafless phase, which include 50 trees.

Table 1: The UAV airborne LiDAR scan data and flight parameters

Related parameters	The parameters settings
Projection mode	UTM
Coordinate system	WGS-84
Relative flight altitude	70 m
Flight speed	6 m/s
Route spacing	45 m
Scan overlap	100%
Scanning frequency	100 kHz

2.3 Data processing flow

In addition to the coordinate information of the target, there is some noise in the original airborne LiDAR point cloud data. It is necessary to remove noise in pre-processing to improve the quality of the data. Second, the point cloud data are filtered and classified to separate the ground point and non-surface points. Then, the DBSCAN, K-means, and the proposed algorithm based on DBSCAN and K-means are used to cluster the pre-processed point cloud data to realize single tree segmentation of vegetation in the mining area. Finally, the segmentation results based on the three clustering algorithms are compared with the measured data and high-resolution images. In this work, we use LiDAR360 software to denoise and filter point cloud data. We use Python programming language for the implementation of the proposed method. The implementation process of the improved algorithm in this paper is shown in Figure 1.

3 The segmentation method of single tree of vegetation in mining area

3.1 Point cloud data preprocessing

In this work, to reduce the quantity of data and improve the segmentation accuracy, it is necessary to remove other types of points and only retain the vegetation points. This work uses LiDAR360 software to denoise the original point cloud. In this software, the ground points are obtained using the filtering algorithm based on irregular triangulation network. The original point cloud data are divided into ground points and non-ground points. The non-ground point data are the point cloud data of the follow-up operation in this work. The comparison chart before and after point cloud filtering is shown in Figure 2.

3.2 DBSCAN clustering algorithm

The DBSCAN clustering algorithm is a classical spatial clustering algorithm based on density. This algorithm starts from randomly selected core point and recursively classifies the points that meet the density requirements into a class. Finally, it obtains the maximized region

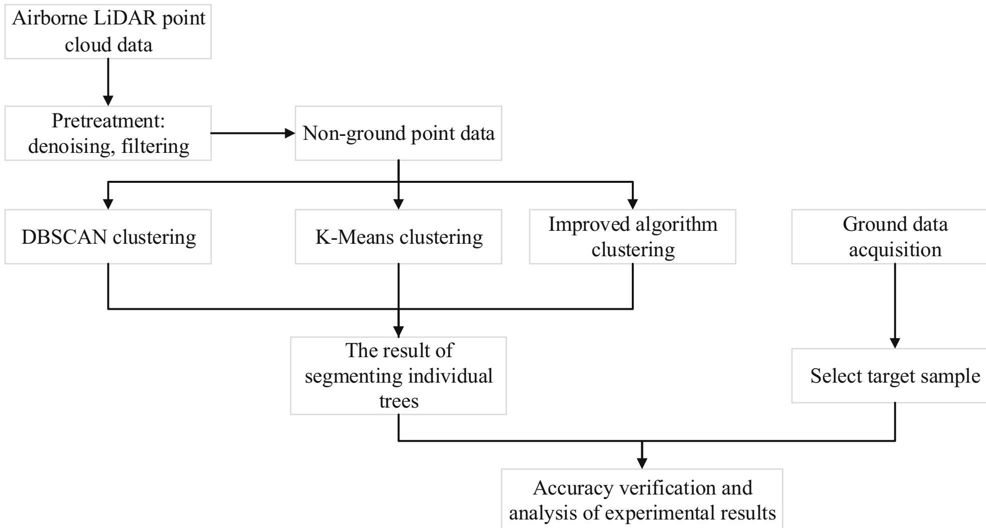

Figure 1: Workflow diagram.

including core points and boundary points. During execution, the DBSCAN algorithm does not require defining the number of clusters, but it needs two parameters: Eps and MinPts. The Eps represents the radius of the cluster, and MinPts denotes the least number of points in the cluster. The core point means that the number of points in the neighborhood of the point is not less than MinPts. The workflow of DBSCAN algorithm is shown in Figure 3.

The major steps of this algorithm are as follows:

Input: 3D point cloud data of the trees, and the initial Eps and MinPts values.

Output: the number of cluster centers and 3D coordinates.

Step 1. Load data and mark all objects as unvisited;

Step 2. Randomly select an untagged object P and mark it as visited. Calculate whether it contains at least MinPts objects in its Eps neighborhood. If so, a new cluster C is established, and P is added to C;

Step 3. Make the collection of all objects in the Eps neighborhood where N is P, select the unmarked objects in N as visited, and calculate the number of objects in its neighborhood. If the number of objects is greater than the MinPTs, put the objects in its neighborhood into the N set. Otherwise, add the object to C;

Step 4. Repeat Step 3 until the objects in the N collection are empty;

Step 5. Repeat Steps 2–4, until all objects are either assigned to a cluster or are marked as noise.

The advantages of DBSCAN algorithm are that it recognizes classes of arbitrary shape, allows clustering into different sizes for each class, does not require defining the number of clusters, and efficiently recognizes noise. However, the computational efficiency of this algorithm is low, and it is unable to deal with high-

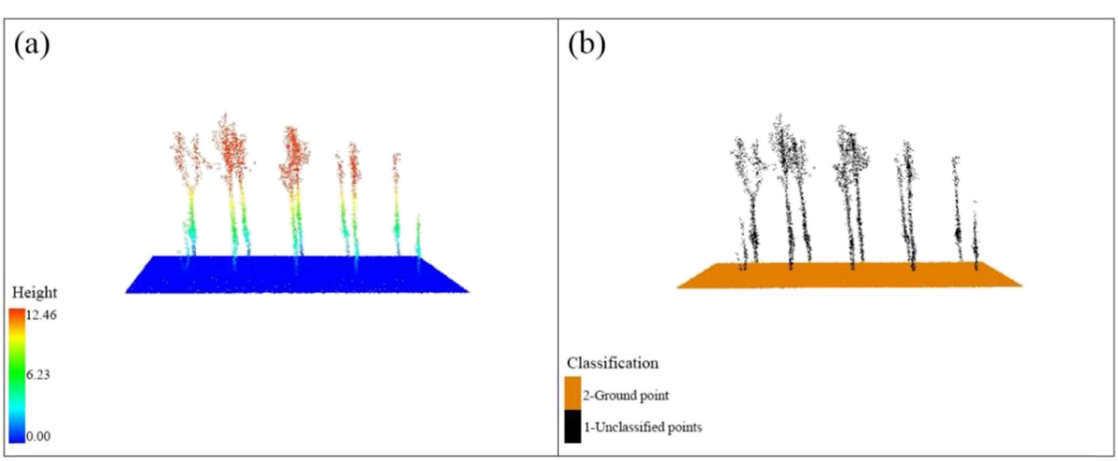

Figure 2: The comparison chart before and after point cloud filtering. (a) Point cloud distribution map before filtering; (b) filtered point cloud distribution map.

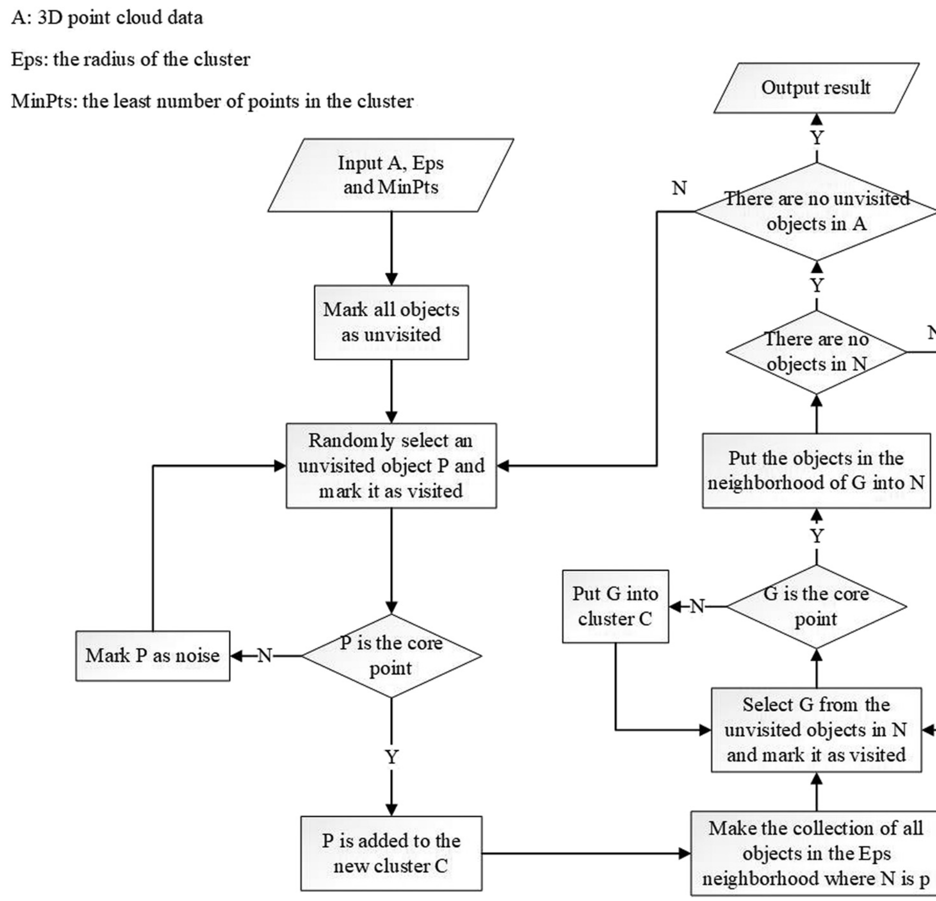

Figure 3: Workflow of DBSCAN algorithm.

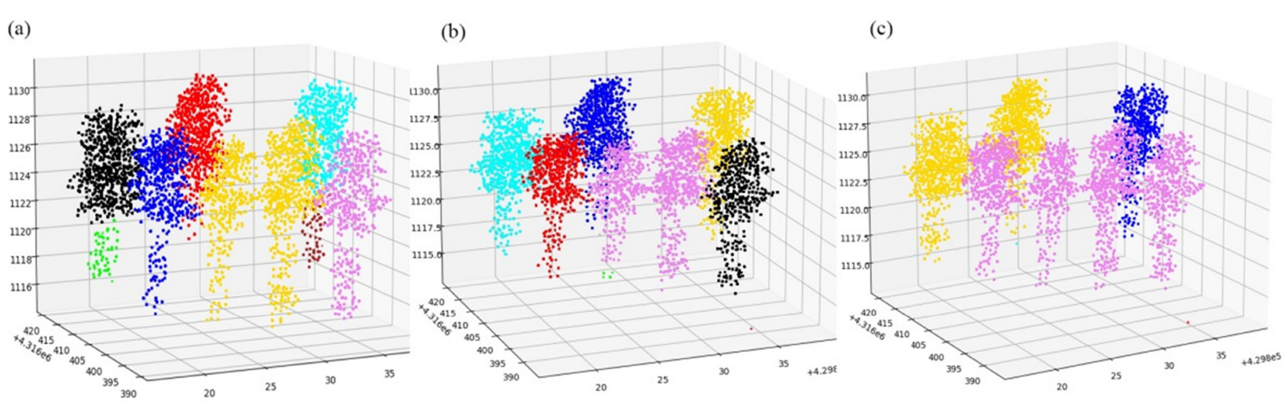

Figure 4: The segmentation results of the DBSCAN algorithm with three different values of Eps and MinPts. (a) Eps = 1, MinPts = 10; (b) Eps = 1, MinPts = 1; (c) Eps = 2, MinPts = 1.

dimensional data effectively. The value of Eps has a great impact on the results. The experimental results show that the segmentation effect of Eps = 1 and MinPts = 1 is better, as shown in Figure 4.

3.3 K-means clustering algorithm

The K-means clustering algorithm is a classic partitioned clustering algorithm. Because of its simplicity computational efficiency, it is one of the well-known clustering algorithms. The K-means clustering algorithm recursively shifts the cluster center coordinates to minimize the functional relationship between each central point and each internal member [22].

The K-mean is mathematically expressed as

$$E = \sum_{j=1}^{k} \sum_{x \in D_j} (x - x_j)^2, \tag{1}$$

where k represents the number of clusters, D_j represents the j class of clustering, x represents any point in D_j, and x_j represents the mean of D_j. E denotes the sum of the square of the distance from the sample point to the cluster center in each class. The smaller the value of E is, the better is the clustering result. The workflow of K-means algorithm is shown in Figure 5.

The major steps of this algorithm are presented below:

Input: The 3D point cloud data of the trees in the mining area, and the number of clusters denoted by k.
Output: The clustering results.
Step 1. Load the data and randomly select k samples as the initial centroids;
Step 2. Calculate the Euclidean distance between each sample and each centroids;
Step 3. Classify each sample into the nearest cluster;
Step 4. Find the average of the samples of each class as the new cluster centroids;
Step 5. Repeat Steps 2–4 until the centroid of the class no longer changes or the number of iterations is reached, and the algorithm ends.

Although K-means algorithm is simple, computationally efficient, and effectively processes the large number of data, there are still some shortcomings. First, it has high requirements for the selection of the initial cluster's mean values. The initial clustering center of the K-means algorithm is randomly selected by the algorithm, and the selection of an initial center represents a result. Especially, when the termination condition of the algorithm is set to stop the recursive computation when the specified number of iterations is reached, the result of the algorithm may not represent the optimal solution.

3.4 Clustering algorithm combining DBSCAN and K-means

In this work, on the basis of in-depth research on DBSCAN clustering algorithm and K-means clustering algorithm, we combine these algorithms with the characteristics of vegetation point cloud data in mining areas and propose an improved single tree segmentation algorithm. Keeping in view the advantages of DBSCAN algorithm, first, the preprocessed data are clustered by DBSCAN to obtain the number and coordinates means of K-means clustering centers. However, the point cloud data of vegetation studied in this work are characterized by wide canopy and thin trunk. In addition, the distance between the crown vertex of the tree and the lowest point of the trunk is large. Now, if the DBSCAN algorithm is directly used to cluster 3D point cloud data, some points may have too small density

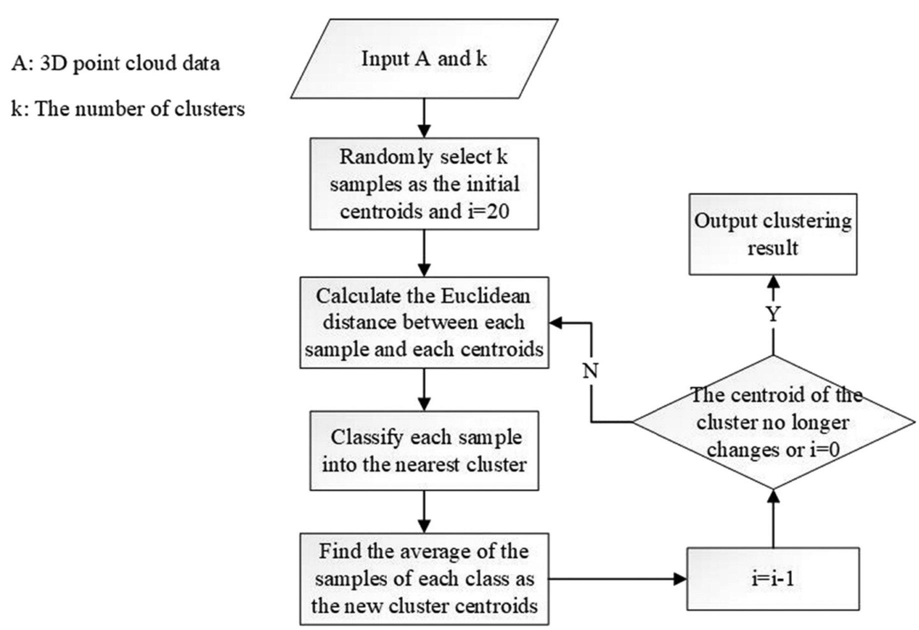

Figure 5: Workflow of K-means algorithm.

because of too large Z value. Therefore, the core point may not be selected accurately. Therefore, in this work, we choose to project the 3D data to the XOY plane, and then process the 2D point cloud data to obtain the number and coordinates of clustering centers. According to the data density of point cloud and the characteristics of trees considered in this work, the experimental results show that the segmentation result of Eps = 2 and MinPts = 14 is best. Then, the point cloud data are clustered based on the classical K-means clustering algorithm. The value of k and the initial clustering center coordinates are obtained by the DBSCAN clustering algorithm. The Euclidean distance is used to calculate the distance from the other points in the data set to the clustering center. The clustering results of 3D point cloud data are finally obtained by following the principle of nearest distance distribution. The workflow of the improved method is shown in Figure 6.

The major steps of the algorithm are as follows:

Input: 3D point cloud data of the trees in the mining area, and the initial values of two parameters, i.e., Eps and MinPts.

Output: The clustering results and the coordinates of the clustering center.

Step 1. Load the data, traverse each sample, delete its Z value, and project it onto the XOY plane;

Step 2. Load the projected data into 2D DBSCAN clustering algorithm;

Step 3. After the 2D DBSCAN clustering algorithm is completed, the Z value is correctly added to each object;

Step 4. Then the average value of the 3D coordinates of each cluster and the number of clusters are calculated and used as the initial value of the K-means clustering algorithm;

Step 5. Finally the original 3D point cloud data are clustered by K-means.

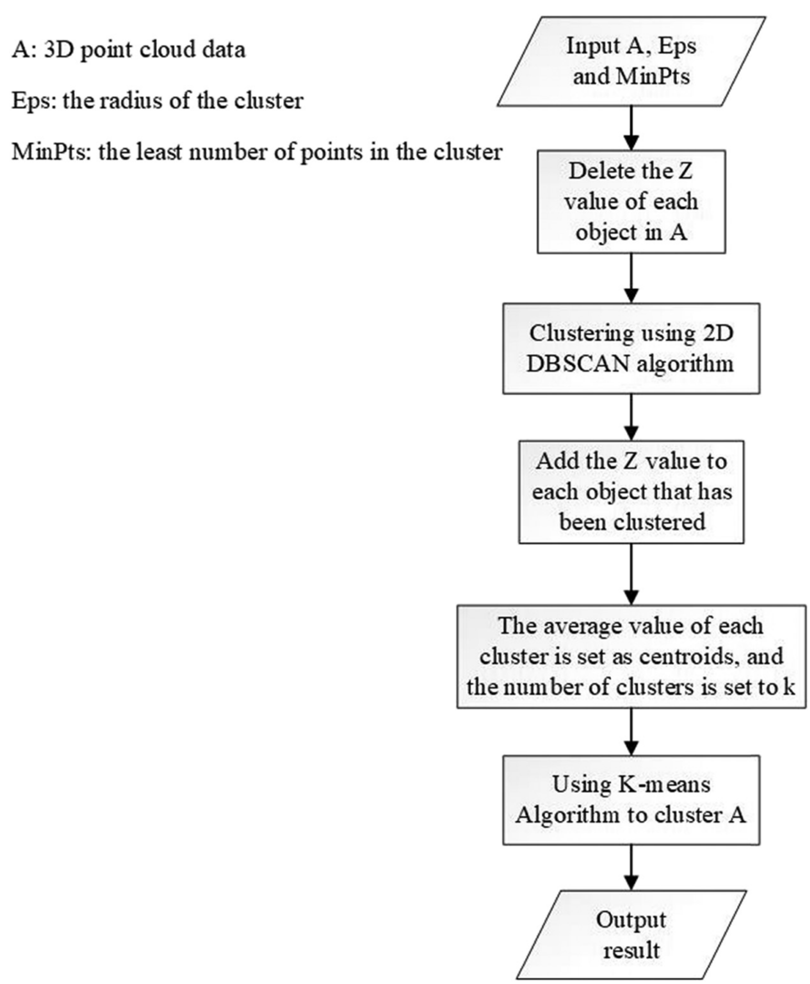

Figure 6: Workflow of the improved method.

4 Accuracy verification and analysis of experimental results

4.1 The accuracy evaluation index of the segmentation algorithm

We assess the validity of these three methods for segmenting individual trees. This is achieved by comparing the results of the algorithm segmentation with the results of the visual interpretation. In this work, we use three evaluation metrics to assess the performance of the aforementioned algorithms. These metrics include recall (r), precision (P), and F1-score (F1) [23]. Furthermore, we also explore the applicability of the three single tree segmentation methods in the vegetation of mining area.

r represents the ratio of the number of effective single trees detected by the algorithm to the actual number of trees. This is mathematically expressed as

$$r = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{2}$$

P represents the ratio of the number of effective single trees detected by the algorithm to all the results extracted as presented in equation (3).

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{3}$$

F1 represents the overall evaluation index of the single tree segmentation accuracy of the algorithm. The larger the F1 score, the higher the segmentation accuracy of the method. This is presented in equation (4).

$$\mathrm{F1} = \frac{2Pr}{P + r}. \tag{4}$$

Please note that the TP represents the number of trees that have been correctly identified, FN represents the number of trees that have not been identified, and FP represents the number of trees that are identified incorrectly.

4.2 Analysis of experimental results

In this paper, we select the segmentation results of a sample plot to show. This intuitively shows the segmentation of individual trees in different phases using three types of tree segmentation methods, which is presented in Figure 7. The high-resolution orthoimages obtained in the same phase is combined with the field survey results as the standard of accuracy evaluation in this work.

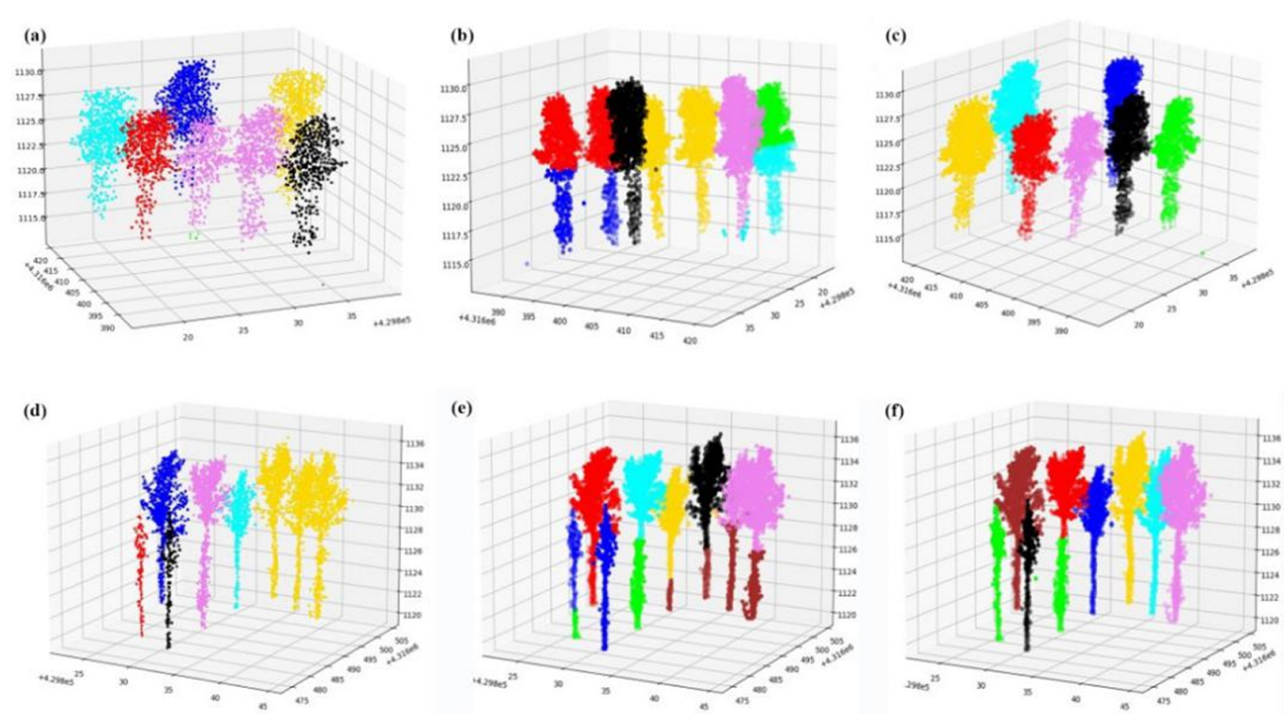

Figure 7: The schematic diagram of tree segmentation in leafy and leafless mining areas using three different clustering algorithms. (a) Leaf stage-DBSCAN, (b) leafl stage-K-means, (c) leafl stage-Improved method, (d) leafless stage-DBSCAN, (e) leafless stage-K-means, and (f) leafless stage-Improved method.

Table 2: The F1-score table of tree segmentation in mining area

Segmentation method	Segmentation accuracy of leafy phase			Segmentation accuracy of leafless phase		
	r	*P*	F1	*r*	*P*	F1
DBSCAN	0.532	1.000	0.694	0.864	0.974	0.916
K-means	0.852	0.535	0.657	0.667	0.432	0.525
Improved method	0.950	0.844	0.894	0.955	0.977	0.966

These are compared with the results of three single tree segmentation methods. The results are statistically analyzed and are presented in Table 2.

As shown in Figure 7, the distribution density of trees in the target area is high, the growth in leafy stage is better, and in leafless stage, it is poor. Based on the traditional DBSCAN method, the single-tree segmentation of the trees in the mining area causes serious omissions. The segmentation results based on the K-means method have some trees that are over-segmented. On the contrary, the improved method obtains a good segmentation of the trees in the mining area. It is also noticeable that there is only a small number of missed classification and misclassification.

As shown in Table 2, the K-means is lower segmentation accuracy for point cloud data in the leafless phase than the leafless phase. The other two methods have a better effect on the individual tree segmentation of point cloud data in the leafless phase, and their F1-scores are greater than 0.90. This is because the canopy structure is complex, and there is an overlap between the canopies in the leafy stage. The point cloud density obtained in the leafy stage is much higher than the leafless stage, which leads to the difficulty of individual tree segmentation at a later stage.

To show the accuracy comparison of results of different algorithms more intuitively, we present the result of Table 2 in Figure 8.

In the segmentation results of the point cloud data of the leafy stage, the value of *r* obtained by the DBSCAN segmentation algorithm is very low, i.e., 0.532. This is because the presence of the leaves make the crown width of the arbor quite different, consequently leading to mutual occlusion. As a result, the density of the obtained point cloud data is not uniform. Now, the DBSCAN segmentation algorithm which is a density-based clustering method does not perform well. The K-means and the proposed segmentation algorithm require the number of clustering centers in advance, which leads to improvement in *r* as compared to the DBSCAN algorithm. The *P* value of the proposed method is much higher than the K-means algorithm. Therefore, we observe that the proposed method has the highest processing accuracy for leafy point cloud data.

When the leaves fall, the airborne laser LiDAR obtains the point cloud information of the tree branches. Because of the absence of leaves, there is no occlusion. In addition, the point cloud density is much smaller as compared to the leafy stage, and the distance between the trees is clearly defined. Therefore, it is notable that the segmentation result of DBSCAN and the proposed algorithm are better as compared to the K-means clustering algorithm. The F1-score difference of the former two algorithms is slightly smaller and higher than 0.9. On the contrary, the F1-score of the K-means algorithm is much lower. This shows that the absence of leaves in the target

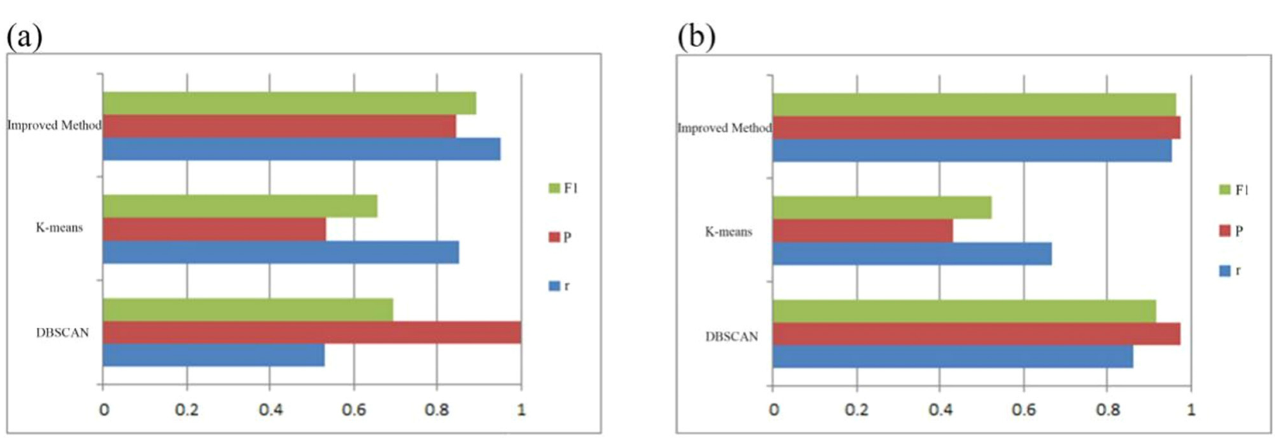

Figure 8: The accuracy of single tree segmentation in leafy and leafless stages. (a) Accuracy of single tree segmentation in leafy stage, and (b) accuracy of single tree segmentation in leafless stage.

area has a great impact on the segmentation results of the DBSCAN algorithm, but not on the results of the other two methods.

In short, although all three methods have the phenomenon of misclassification, the proposed method shows effective results in both phases of the two-phase experiments. On the contrary, the DBSCAN segmentation algorithm is greatly influenced by the presence or absence of tree leaves. The results of K-means algorithm in the two-phase experiments are the worst. Therefore, it is evident that the proposed method for single tree segmentation of vegetation in mining areas is an effective technique that outperforms other methods.

5 Discussions

In this paper, the improved algorithm is used to segment the point cloud data of leafy and leafless stage in the study area, and the segmentation results are compared with those of traditional K-means and DBSCAN clustering algorithms. The simulation results and analysis show that the single tree segmentation method proposed in this work has high accuracy and good efficiency.

Since the emergence of LiDAR, it has shown great potential in the field of forest science [24–27]. At present, the study of vegetation in mining area can strengthen the management of ecological environment in mining area, which is of great significance for the protection of ecological environment. However, now people need to study the forest more carefully. The successful identification and characterization of individual trees are critical in forest science.

However, it is more difficult to segment some trees with complex structures [28]. In some past studies, the accuracy of most individual tree segmentation can only reach 50–80% [15,29]. For example, Koch et al. divided the canopy of coniferous forest and broad-leaved forest, and the overall accuracy rate was 62% [14]. Wang et al. overlaid the automatically detected tree-crown image with one of the manually delineated crown images and found that the accuracy rate was 75.6% [13]. Kwak et al. reported accuracies of their new algorithm from 60 to 80% in coniferous and deciduous forests [30]. In this paper, the new algorithm is directly applied to point cloud data, and the accuracy of segmenting individual trees has been improved. The accuracy of the leafless stage is 98%, and the accuracy of the leaf stage can also reach 84%. The new algorithm shows an excellent segmentation effect for trees with simple structure, but it is still poor for trees with complex structures, especially for trees with small intervals.

When choosing a single tree segmentation algorithm, we must consider not only the accuracy of the algorithm, but also the efficiency of the algorithm. Many segmenting individual trees algorithms use the normalized cut segmentation method, but the computational complexity of the normalized segmentation is too large, which will lead to low efficiency. For example, the segmenting individual trees algorithms proposed by Reitberger et al. and Yao et al. based on normalized cut segmentation do not have high segmentation efficiency [18,31]. Even if Yao et al. improves the previous algorithm and proposes the combination of normalized cut segmentation method and DBSCAN, the segmentation efficiency has not been greatly improved [32]. Kandare et al. use 2D and 3D K-means clustering algorithm to segment individual trees above and under the forest respectively. The segmentation efficiency of this algorithm is high, but it requires high initial value and threshold setting [20].

In this paper, the improved algorithm can not only improve the efficiency of single tree segmentation, but also obtain better results. Because the research object of this paper has the characteristics of wide canopy and thin trunk, if we directly use the three-dimensional DBSCAN clustering algorithm for segmenting individual trees, it will be impossible to accurately select the core point. Therefore, the improved algorithm projects the original three-dimensional point cloud data to the XOY plane, and then uses the DBSCAN clustering algorithm to segment. The number and coordinates of clustering centers can be calculated and input to the K-means algorithm as an initial value. Finally, we can get a more accurate result of segmenting individual trees. This algorithm can greatly accelerate the processing speed of DBSCAN clustering algorithm and can provide a more accurate initial value for K-means clustering algorithm, which is of great significance for the final segmentation results.

In our algorithm, the uncertainty in tree segmentations mainly derives from the size of the threshold. In the forest where the tree spacing is large, we can choose a relatively large threshold to segment individual trees. However, it is difficult to select the threshold. A higher threshold will lead to under-segmentations whereas a smaller threshold can lead to over-segmentations. Therefore, in future research, we can consider using an adaptive threshold to segment individual trees. The improved algorithm has not been tested on mixed forests. The applicability of the algorithm to other types of forests is not clear. In future research, we can test forests with mixed forests.

6 Conclusion

This paper proposes an improved algorithm of segmenting individual trees based on LiDAR point cloud data. The algorithm is based on the advantages of DBSCA and K-means clustering algorithms. The experimental results show that the proposed algorithm has higher accuracy in both the leafy stage and the leafless stage. The accuracy, in terms of Recall, Precision, and F1-score, is relatively high, indicating that the improved algorithm has good potential for use in other forested areas. Through the individual tree segmentation of the vegetation in the mining area, the researchers can accurately extract various parameters of the vegetation. We compare and analyze the vegetation parameters in the early and later stages of green mine construction. This can quantitatively evaluate the construction of green mines in the experimental area and provide effective references for relevant departments to carry out environmental protection and construction in mining areas in the future. We can also monitor the vegetation parameters to clarify the impact of mining activities on vegetation in the mining area and provide a direction for the construction of green mines.

Funding information: This work was financially supported by the Research Project of Key Technologies for Water Resources Protection, Utilization, and Ecological Reconstruction in Northern Shaanxi Coal Mine Area (SMHKJ-A-J-03:2018).

Author contributions: Conceptualization and methodology, Qiuji Chen and Xin Wang; investigation, Mengru Hang and Jiye Li; writing—original draft preparation, Xin Wang; writing—review and editing, Qiuji Chen. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

[1] Li BJ, Gu HH, Ji YZ. Evaluation of landscape pattern changes and ecological effects in land reclamation project of mining area. Trans CSAE. 2012;28(3):251–6. doi: 10.3969/j.issn.1002-6819.2012.03.043.

[2] Chen J, Jiskani IM, Jinliang C, Yan H. Evaluation and future framework of green mine construction in China based on the DPSIR model. Sustain Environ Res. 2020;30(1):1–10. doi: 10.1186/s42834-020-00054-8.

[3] Zhou Y, Zhou W, Lu X, Jiskani IM, Cai QX, Liu P, et al. Evaluation Index System of Green Surface Mining in China. Mining Metall Explor. 2020;37:1093–103. doi: 10.1007/s42461-020-00236-3.

[4] Jiskani IM, Cai QX, Zhou W, Shah SAA. Green and climate-smart mining: a framework to analyze open-pit mines for cleaner mineral production. Resour Policy. 2021;71:102007. doi: 10.1016/j.resourpol.2021.102007.

[5] Ma CG, Jiang W. How to protect the ecological environment in the green mine demonstration area. Resour Econ Environ Prot. 2020;227(10):47–8. doi: 10.16317/j.cnki.12-1377/x2020.10.018.

[6] Huang Y, Wang YJ, Tian F, Tan XL. NPP lost in mining area based on improved FPAR. J Arid Land Resour Environ. 2015;29(2):49–54. doi: 10.13448/j.cnki.jalre.2015.045.

[7] Li ZY, Liu QW, Pang Y. Review on forest parameters inversion using LiDAR. J Remote Sens. 2016;20(5):1138–50. doi: 10.11834/jrs.20165130.

[8] Zhang CY, Zhou YH, Qiu F. Individual tree segmentation from LiDAR point clouds for urban forest inventory. Remote Sens. 2015;7(6):7892–913. doi: 10.3390/rs70607892.

[9] Wulder M. Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. Prog Phys Geogr. 1998;22(4):449. doi: 10.1191/030913398675385488.

[10] Liu LX, Pang Y. Applications of airborne laser scanning and terrestrial laser scanning to forestry. World Forestry Res. 2014;27(1):49–56. doi: 10.13348/j.cnki.sjlyyj.2014.01.009.

[11] Zhang WJ. 3D Laser scanning technology and its application. Stand Surv Mapp. 2016;32(2):42–4. doi: CNKI:SUN:CHBC.0.2016-02-018.

[12] Hyyppa J, Kelle O, Lehikoinen M, Inkinen M. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. IEEE Trans Geosci Remote Sens. 2001;39(5):969–75. doi: 10.1109/36.921414.

[13] Wang L, Gong P, Biging GS. Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery. Photogramm Eng Remote Sens. 2004;70(3):351–8. doi: 10.14358/PERS.70.3.35.

[14] Koch B, Heyder U, Weinacker H. Detection of individual tree crowns in airborne LiDAR data. Photogramm Eng Remote Sens. 2006;72(4):357–63. doi: 10.14358/PERS.72.4.357.

[15] Chen Q, Baldocchi D, Gong P, Kelly M. Isolating individual trees in a Savanna woodland using small footprint LiDAR data. Photogramme Eng Remote Sens. 2006;72(8):923–32. doi: info:doi/10.14358/PERS.72.8.923.

[16] Mohan M, Araujo B, Mendona FD, Silva CA, Cardil A. Optimizing individual tree detection accuracy and measuring forest uniformity in coconut (Cocos nucifera L.) plantations using airborne laser scanning. Ecol Model. 2019;409:108736. doi: 10.1016/j.ecolmodel.2019.108736.

[17] Guo Q, Li W, Yu H, Alvarez O. Effects of topographic variability and LiDAR sampling density on several DEM interpolation methods. Photogramm Eng Remote Sens. 2010;76(6):701–12. doi: 10.14358/PERS.76.6.701.

[18] Reitberger J, Schnoerr C, Krzystek P, Stilla U. 3D segmentation of single trees exploiting full waveform LiDAR data. ISPRS J

Photogramm Remote Sens. 2009;64(6):561–74. doi: 10.1016/j.isprsjprs.2009.04.002.
[19] Ayrey E, Fraver S, Kershaw J, Kenefic L, Hayes D, Weiskittel A. Layer stacking: a novel algorithm for individual forest tree segmentation from LiDAR point clouds. Can J Remote Sens. 2017;43(1):16–27. doi: 10.1080/07038992.2017.1252907.
[20] Kandare K, Da Lponte M, Gianelle D, Chan CW. A new procedure for identifying single trees in understory layer using discrete LiDAR data. Geoscience and Remote Sensing Symposium. New York: Institute of Electrical and Electronics Engineers Inc; 2014. doi: 10.1109/IGARSS.2014.6946686.
[21] Li W, Guo Q, Jakubowski MK, Kelly M. A new method for segmenting individual trees from the LiDAR point cloud. Photogramm Eng Remote Sens. 2012;78:75–84. doi: info:doi/10.14358/PERS.78.1.75.
[22] Guo JZ, Zhang M, Fei ZX, Zhang LL, Jiang XD, Yang M. Principal component analysis and cluster analysis of quantitative characters of Xanthoceras sorbifolia. J Northwest Forestry Univ. 2012;27(2):66–9. doi: 10.3969/j.issn.1001-7461.2012.02.14.
[23] Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. Comput Sci. 2005;3408:345–59. doi: 10.1007/978-3-540-31865-1_25.
[24] Dubayah RO, Drake JB. LiDAR remote sensing for forestry. J Forestry. 2000;98(6):44–6. doi: 10.1007/BF02769134.
[25] Bjerknes NO. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. Remote Sens Environ. 2001;78(3):328–40. doi: 10.1016/S0034-4257(01)00228-0.
[26] Lim K, Treitz P, Wulder M, St-Onge B, Flood M. LiDAR remote sensing of forest structure. Prog Phys Geogr. 2003;27(1):88–106. doi: 10.1191/0309133303pp360ra.
[27] Popescu SC, Wynne RH. Seeing the trees in the forest: using LiDAR and multispectral data fusion with local filtering and variable window size for estimating tree height. Photogramm Eng Remote Sens. 2004;70(5):589–604. doi: 10.14358/PERS.70.5.589.
[28] Holmgren J. Detecting and measuring individual trees using an airborne laser scanner. Photogramm Eng Remote Sens. 2002;23(18):3893–8. doi: 10.1080/01431160210137695.
[29] Alexander C. Delineating tree crowns from airborne laser scanning point cloud data using Delaunay triangulation. Int J Remote Sens. 2009;30(13–14):3843–8. doi: 10.1080/01431160902842318.
[30] Kwak DA, Lee WK, Lee JH, Biging GS, Gong P. Detection of individual trees and estimation of tree height using LiDAR data. J For Res. 2007;12(6):425–34. doi: 10.1007/s10310-007-0041-9.
[31] Yao W, Krzystek P, Heurich M. Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. Remote Sens Environ. 2012;123:368–80. doi: 10.1016/j.rse.2012.03.027.
[32] Yao W, Krzystek P, Heurich M. Enhanced detection of 3D individual trees in forested areas using airborne full-waveform LiDAR data by combining normalized cuts with spatial density clustering. Biopolymers. 2013;II-5/W2:349–54. doi: 10.5194/isprsannals-II-5-W2-349-2013.