Geodesy and Geodynamics 13 (2022) 38–49

Contents lists available at ScienceDirect

Geodesy and Geodynamics

journal homepage: http://www.keaipublishing.com/geog

A novel fast classification filtering algorithm for LiDAR point clouds based on small grid density clustering

Xingsheng Deng*, Guo Tang, Qingyang Wang

School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha, 410114, China

ARTICLE INFO

Article history:
Received 11 August 2021
Accepted 27 October 2021
Available online 11 November 2021

Keywords:
Small grid density clustering
DBSCAN
Fast classification filtering algorithm

ABSTRACT

Clustering filtering is usually a practical method for light detection and ranging (LiDAR) point clouds filtering according to their characteristic attributes. However, the amount of point cloud data is extremely large in practice, making it impossible to cluster point clouds data directly, and the filtering error is also too large. Moreover, many existing filtering algorithms have poor classification results in discontinuous terrain. This article proposes a new fast classification filtering algorithm based on density clustering, which can solve the problem of point clouds classification in discontinuous terrain. Based on the spatial density of LiDAR point clouds, also the features of the ground object point clouds and the terrain point clouds, the point clouds are clustered firstly by their elevations, and then the plane point clouds are selected. Thus the number of samples and feature dimensions of data are reduced. Using the DBSCAN clustering filtering method, the original point clouds are finally divided into noise point clouds, ground object point clouds, and terrain point clouds. The experiment uses 15 sets of data samples provided by the International Society for Photogrammetry and Remote Sensing (ISPRS), and the results of the proposed algorithm are compared with the other eight classical filtering algorithms. Quantitative and qualitative analysis shows that the proposed algorithm has good applicability in urban areas and rural areas, and is significantly better than other classic filtering algorithms in discontinuous terrain, with a total error of about 10%. The results show that the proposed method is feasible and can be used in different terrains.

1. Introduction

Light detection and ranging (LiDAR) technology can obtain dense point clouds with three-dimensional coordinates actively and quickly, also obtain some attributes on the surface of an object, which provides help for collecting elevation information in areas that are difficult for people to reach. The airborne laser radar system is currently the most advanced aviation remote sensing system that can obtain three-dimensional information and images of terrain surfaces in real time. The airborne laser scanning system obtains the three-dimensional ground information from various ground objects, so it is necessary to separate the points on the terrain surface and the non-terrain surface [1–3].

The process of removing non-ground points from the point cloud data to obtain a true digital terrain model is often called "non-ground point cloud filtering" [4]. There are many point clouds filtering algorithms, which generally include four types: slope based filtering algorithm, block minimum based filtering algorithm, surface based filtering algorithm, and clustering or segmentation based filtering algorithm [5]. The filtering method based on slope change assumes firstly that the steepest slope belongs to the ground object class, and the point with a slope greater than a certain threshold value is considered to be object point, while the point with a slope less than a certain threshold value is considered to be the ground point. Vosselman clearly defined the acceptable height difference between two points as a function of the distance

* Corresponding author. School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha, China.
E-mail address: whudxs@163.com (X. Deng).

Peer review under responsibility of Institute of Seismology, China Earthquake Administration.

Production and Hosting by Elsevier on behalf of KeAi

https://doi.org/10.1016/j.geog.2021.10.002

between points, and used a special kernel function to filter out ground object points according to the slope threshold [6], which was more suitable for gentle terrain. Based on the above ideas, Sithole adapts to steep terrain with threshold value changing along with terrain slope [7]. The improvement of such method is mostly reflected in the setting of the slope threshold value. For example, the method proposed by Ding Shaopeng includes local slope statistics to obtain local terrain slope values [8]. The minimum block based filtering method sets up a horizontal discriminant function on which the buffer defines the expected position of ground points. Wack and Wimmer use gradient information to replace the points exceeding the height difference threshold with the actual approximate lowest point, and use the Gaussian Laplacian Algorithm to detect and remove non-ground points in turn [9]. Miao Qiguang predicted terrain slope parameters by using elevation standard deviation, and then determined ground points by morphological filtering [10]. The surface-based filtering method defines the discriminant function of the corresponding buffer as a parameterized surface. Elmqvist uses a dynamic contour model to obtain ground points by minimizing the energy function of the model [11]. Liu X. obtained the virtual ground seed points from the set virtual seed points through the multi-scale morphology method. After the initial TIN was generated, the filtering was completed through iteration [12]. If any point of the cluster is higher than its neighbors, then any point of the cluster should be a non-ground point. Otherwise it should be a bare ground point. The above assumption is the rationality of the filtering method based on clustering or segmentation [13]. George Sithole et al. proposed a method to divide the point clouds into smooth segments that still contain high discontinuities, and classify the resulting segments by comparing the topological and geometric relationships with adjacent segments [14]. Nesrine Chehata et al. proposed the method to use the K-means algorithm to filter the point clouds hierarchically, and then use the local slope map to improve the filtering of ground points [15].

Based on the analysis of existing clustering filtering methods, from the perspective of improving algorithm automation and making full use of the original point clouds information, by using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [16], this article proposes a fast classification filtering algorithm based on density clustering. The algorithm only uses the information of one plane of the point clouds during the initial clustering, which greatly reduces the complexity of the calculation. It fully considers the height difference relationship between the neighborhoods of the point clouds. After multiple clustering, the point clouds are classified into the non-ground class and the ground class, to achieve the purpose of removing the non-ground points from the total point clouds.

2. DBSCAN algorithm

Cluster analysis is an important research direction in data mining and can be used as a preprocessing step for other analysis algorithms. It divides the data into different and meaningful subclass, making the data of the same subclass as similar as possible and the data of different subclass as different as possible [17–19]. Clustering algorithms have a variety of classification methods [20], which can be roughly divided into hierarchical clustering algorithms [21,22], partitioned clustering algorithms [23,24], density-based, and gird-based clustering algorithms [25–28], and other clustering algorithms [29].

DBSCAN is a density-based spatial clustering algorithm, which is different from the convex clusters of hierarchical clustering algorithm and partitioned clustering algorithm. It uses the density connectivity of classes to divide regions with sufficient density into clusters. It can quickly find clusters of any shape in a noisy spatial database, without pre-determining the number of classes, and has the ability to handle large databases. Assuming a given data set $D = x_i$, the basic definitions involved in the DBSCAN algorithm are as follows [30,31]:

Definition 1(Neighborhood) For $x_j \in D$, Epsilon is the radius of a search area, the neighborhood contains the samples in D whose distance from x_j is less than or equal to Epsilon, that is:

$$NEps(x_j) = \{x_j \in D | Dist(x_i, x_j) \leq Eps\} \tag{1}$$

$Dist(x_i, x_j)$ is the distance between objects x_i and x_j; $NEps(x_j)$ is all objects in D whose distance from x_j is not greater than Epsilon.

Definition 2(Core Object) Set the density threshold of the neighborhood to MinPts, if the neighborhood of x_j contains at least MinPts samples, that is:

$$|NEps(x_j)| \geq MinPts \tag{2}$$

when x_j is the center of the circle, there are at least MinPts objects in D within the radius, then x_j is said to be a core object.

Definition 3(Directly Density-reachable) For x_i and x_j, if x_j is in the neighborhood of x_i and x_i is the core object, then it is said that x_i is directly density-reachable from x_i.

Definition 4(Density-reachable) If there is an object chain x_1, x_2, ..., $x_n \in D$, for $x_i (0 < i < n)$, x_{i+1} is directly density-reachable from the object x_i, then the object x_n is density-reachable from the object x_1, and the density-reachability is asymmetric.

Definition 5(Density-connected) For x_i and x_j, if x_k exists so that both x_i and x_j can be density-reachable from the object x_k, then x_i and x_j are density-connected, and the density connection has symmetry.

Definition 6(Clusters and noise) Take any point p from D, start from point p, search for all points in D that satisfy the "neighborhood" parameter (Epsilon, MinPts) and the density is reachable, then these points constitute cluster C. If p does not belong to any cluster, it is marked as a noise point:

$$noise = \{p \in D | \forall i : p \notin C_i\} \tag{3}$$

Equation (3) represents the set of noise points, where C_i represents the cluster in the data set D.

The above definition can be explained in Fig. 1. When $MinPts = 3$, the red dot is the core point, the yellow dot is the boundary point, and the blue dot is the noise point. B is directly density-reachable

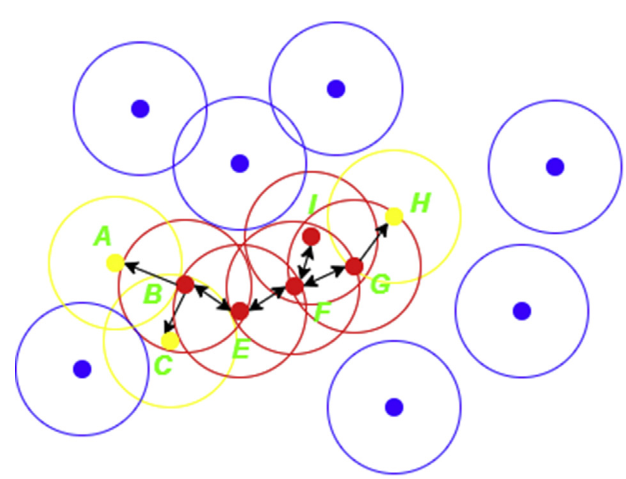

Fig. 1. Schematic diagram of DBSCAN.

from E, and A is directly density-reachable from B, so A is density-reachable from E, but E cannot reach the density from A (asymmetry). A、B、C、E、F、G、H and I are all density-connected (symmetry).

The basic process of DBSCAN clustering [19,30]:marking all samples in D as unprocessed, setting parameters Epsilon and MinPts, finding all the objects which are density-reachable from the unprocessed sample. If the sample point is the core point, it means that the number of objects meeting the condition in the neighborhood is not less than MinPts. Thus continue to expand with this sample point as the center, and finally, find a class corresponding to parameters Epsilon and MinPts. If the number of objects meeting the condition in the neighborhood is less than MinPts, the sample points are temporarily marked as noise points, and the next object is processed. According to whether the sample point is in the neighborhood of the core point, it is judged whether it is a noise point, and if it is in the neighborhood of the core point, it is marked as a boundary point.

3. Fast classification filtering algorithm based on density clustering

The basic ideas of the proposed fast classification filtering algorithm based on density clustering are as follows: for a set of point clouds data, a single point has its three-dimensional coordinate information, and multiple points can determine the height difference and horizontal distance between points. In most areas, ground objects such as buildings and trees, terrain such as flat land and hillsides, their point clouds will show different characteristics in space. At the boundary of the ground object point clouds, there will be a large height jump within a small horizontal distance; while inside the point clouds, the height difference changes little. The overall terrain point clouds change smoothly in space. Such characteristics are reflected in the density change in the space. The junction of the ground object and the terrain belongs to the low-density area of this area, while the interior of the ground object and the terrain belongs to the high-density area of this area, and the high density set is separated by the low density set. This article believes that the density change characteristics of continuous density in the same class but discontinuous density between different classes are very suitable for the use of the DBSCAN algorithm. With only two neighborhood parameters, i.e., Epsilon and MinPts, it is possible to discover high density sets of any shape under unsupervised conditions, thereby dividing the ground object and terrain into two different parts, and there is no need to predetermine the number of clusters, and the recognition is highly automated. For areas with steep ridges, the difficulty of filtering usually lies in the sharp change in slope from gentle terrain to steep ridge terrain, which is difficult to identify at one time. However, this article believes that there are "density-connected" point clouds between the gentle terrain and the steep terrain, so that the steep terrain can be directly connected to the gentle terrain through clustering.

However, in the actual situation of engineering practice, because the amount of point clouds data is extremely large, in direct clustering, the time complexity of the DBSCAN clustering algorithm is $O(n^2)$, where n is the total number of samples. Using DBSCAN clustering directly will lead to too long time for data processing and require large memory support. In the end, not only the clustering results are poor, but clustering may not be implemented at all. In the past few years, many scholars have been trying to solve the dimensional disaster problem of the DBSCAN algorithm. Some improved DBSCAN algorithms have been proposed recently, such as ρ-approximate DBSCAN [32], NG-DBSCAN [33], KNN-DBSCAN [34], KNN-BLOCK DBSCAN [35], BLOCK DBSCAN [36]. These algorithms have made great improvements to make DBSCAN clustering in high-dimensional data more efficient. This article focuses more on the applicability of the DBSCAN algorithm in LiDAR point cloud filtering, but not on the DBSCAN algorithm itself. Considering actual point cloud data characteristics, a fast classification filtering algorithm based on density clustering is proposed. The three main ideas of the proposed algorithm are as follows:

(1) Clustering in lower dimensions and grids. Since the number of point clouds in each grid is much smaller than the total, n in the time complexity $O(n^2)$ will be greatly reduced. For quadratic complexity algorithms, the reduction of n will speed up the operation of the algorithm.
(2) The projections of the ground object and terrain point clouds on the Z axis have obvious density differentiation. The ground object point clouds will produce a clear low-density zone with the terrain point clouds due to the height jump. However, the changes within the ground object point clouds and the terrain point clouds are more continuous and uniform, so it is recognized as a high-density zone.
(3) The ground object point clouds will have obvious height jumps at its boundary, which is specifically manifested as a large height difference change within a short horizontal distance.

The flow of the proposed filtering algorithm is shown in Fig. 2:

The steps of the fast classification filtering algorithm based on density clustering are as follows:

(1) The point clouds are divided into different small grids according to the plane coordinates firstly. In a large survey area, there are topographic clouds and ground objects clouds. In the case of a large number of point clouds, a directly clustering the point clouds according to the elevation value will lead to poor clustering results. However, it can fully retain the features of the ground objects by dividing the large survey area into several small survey areas with a length D firstly.
(2) DBSCAN algorithm is performed to obtain an initial clustering of point clouds by every small grid. Taking the height of the point as the characteristic attribute, the point clouds are divided into several non-attribute classes according to

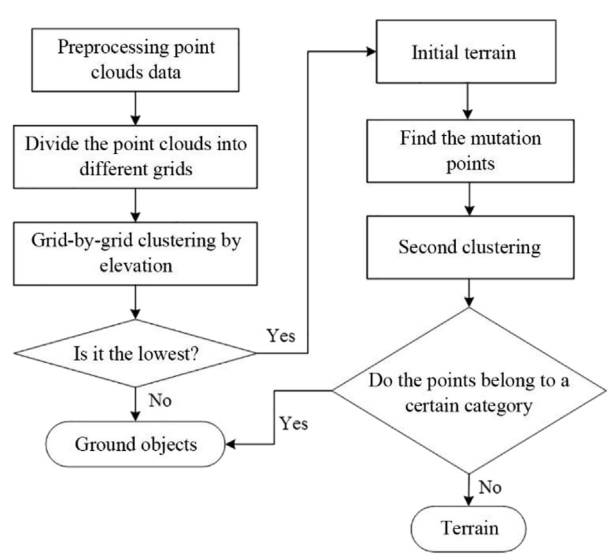

Fig. 2. Flow chart of fast classification filtering algorithm based on density clustering.

parameters Epsilon 1 and MinPts. MinPts is generally considered to be greater than the dimension of the data. Because noise has the characteristics of no core point and low density, clustering will divide the noise into a separate category, to achieve an effect of de-noising.

(3) The appropriate class is chosen to form the initial terrain point clouds. Generally, the lowest class in the surveying area is regarded as the terrain class, and all initial clustering results are summarized to form the initial terrain category. There is no absolute difference in height between terrain and ground objects, and height is only one of the basis for judgment; therefore the initial terrain at this time may still contain ground object points.

(4) The fourth step is to search for the nearest point on the horizontal plane and calculate the height difference. Generally speaking, a common characteristic between the ground object point clouds and the terrain point clouds is that within the same class, the height change is small when the horizontal distance between points is short. The difference is that the boundary point of the ground object will have a large height change within a short horizontal distance. Using range 1 and range 2 as the thresholds, if there is a pair of points, the height difference within the horizontal distance range1 is not less than range 2, which means the higher point can be considered the boundary point of the ground object. As long as some boundary points of a ground object are found, the entire ground object can be found through clustering.

(5) The secondary clustering is performed by using the found boundary points called mutation points. The mutation points are used as initial points to cluster in the initial terrain category according to the threshold (Epsilon 2, MinPts).

(6) The final terrain class is formed lastly. After the clustering result is obtained, if the point belongs to a certain category, the point is considered an object point; otherwise, it is a ground point.

4. Experiment and analysis

In order to validate the filtering performance of the proposed algorithm on the LiDAR point cloud, the LiDAR data collected by ISPRS using the Optech ALTM scanner in the Vaihingen/Enz test area and the center of Stuttgart are selected as the experimental data [37]. Fifteen sample data are selected from the data set with different characteristics, from four areas with urban characteristics and four areas with rural characteristics (Site 1 to 8). The sample data are manually classified into ground points and non-ground points. The terrain features reflected in this data set are obvious and very representative, covering various types of terrain conditions common to filtering, and professional reference results are provided by ISPRS. The results of filtering experiments conducted on this basis are more reliable. And many classic filtering algorithms rely on this data set to make experiments, so enough comparison results are produced. The above is also an important reason for using this data set in this article. Table 1 shows the basic information of fifteen sample data. The experiment will analyze the filtering results of fifteen sample data from qualitative and quantitative angles.

4.1. Quantitative analysis

The cross-matrices of errors are used as evaluation criteria for filtering error to determine the proportion of misclassified ground points and non-ground points [5,37,38]. Table 2 defines the cross-matrices of errors.

Among them, *a* is the number of ground points that are correctly identified as ground points; *b* is the number of ground points that are incorrectly identified as non-ground points; *c* is the number of non-ground points that are incorrectly identified as ground points; and *d* is the number of non-ground points that are correctly identified as non-ground points. Type I error is the rejection error, which refers to the wrong classification of ground points as non-ground points. The smaller the TI is, the more precise ground details can be retained; Type II error is acceptance error, which refers to the wrong classification of non-ground points as ground points. The smaller the TII, the more residual ground object points can be removed. The smaller the total error, the higher the classification quality.

Based on the above error definition, Table 3 shows the filtering error obtained after filtering fifteen samples with the proposed algorithm. In addition, Table 4 compares the fifteen sample filtering errors of other classic filtering algorithms [39,40].

It can be seen from Table 3 that the proposed algorithm can get better results in S21, S42, S54, and S61, and Type I error and total error are small. The total error in other samples is mostly about 10%. Among them, S21 and S42 belong to urban areas, and S54 and S61 belong to rural areas. In Table 4, for S21, S23, S24, S53, S54, S61, and S71, the total error of the proposed algorithm is lower than the average. Among them, S53, S54, S61, and S71 belong to rural areas, and the terrain features are discontinuous terrain, villages, and bridges. The terrain features of S21, S23, and S24 in urban areas are narrow bridges, complex buildings, discontinuous terrain, steep slopes, and vegetation. In general, the proposed algorithm has a certain stability, and the results in urban areas and rural areas are at the same level. It can identify complex buildings of arbitrary shapes and has good results in discontinuous terrain areas.

4.2. Qualitative analysis

In order to analyze the filtering results more specifically, the experiment selected 2 urban areas and 2 rural areas with multiple characteristics, including buildings, bridges, discontinuous terrain, and slopes with vegetation for qualitative analysis.

4.2.1. Experimental sample 21

Sample 21 is an urban sample with narrow bridge characteristics. After filtering, the results of this sample are shown in Fig. 3. White dots represent points that are finally classified as ground points; black dots represent points that are finally classified as object points. In Fig. 4, yellow dots represent ground points that are finally misclassified as object points, and red dots represent object points that are finally misclassified as ground points. In Fig. 5, the blue dots are the points where the ground points are re-divided into ground object points after the second clustering, including the points where the green dots in Fig. 6 are located, and the green dots represent the selected mutation points. Fig. 7 is a shadow relief map of unfiltered data generated by Kriging interpolation, and Fig. 8 is a shadow relief map generated by filtered data.

The finally selected ground object points can conform to the outline of the real ground objects, and the bridge can be accurately separated. The selected mutation points are distributed at the boundary of the ground objects, and through secondary clustering, some of the misclassified ground objects are corrected without error. The bridge in Fig. 8 has been basically removed, and most of the ground objects in the survey area are removed accurately, but it is still affected by other unfiltered points, for example, vegetation points and building points. The filtering error of the sample is mainly type II error, and the remaining part of the bridge in Fig. 8 is caused by type II error. Take this as an example, in the grid where the remaining bridges are located, Fig. 9 shows the distribution of bridge points in elevation, and Fig. 10 shows the classification results of bridge points.

Table 1
Statistics and related description of sample data.

Location	Test data	Terrain feature	Sample	Number of points
City	Site 1	Steep slopes, vegetation, buildings	Sample 11	38,010
			Sample 12	52,119
	Site 2	Highway and bridge, large buildings, irregular buildings	Sample 21	12,960
			Sample 22	32,706
			Sample 23	25,095
			Sample 24	7492
	Site 3	Dense buildings, vegetation	Sample 31	28,862
	Site 4	Railway station, train	Sample 41	11,231
			Sample 42	42,470
Rural area	Site 5	Steep slopes, vegetation, quarries, river banks	Sample 51	17,845
			Sample 52	22,474
			Sample 53	34,378
			Sample 54	8608
	Site 6	Large buildings, road embankment	Sample 61	35,060
	Site 7	Bridge, underground passage	Sample 71	15,645
	Site 8	High bridge, river bank, vegetation	Non	Non

Table 2
Definition of filtering error.

Reference point	Filtered point		Metrics of quantitative evaluations (%)		
	ground points	non-ground points	Type I (TI)	Type II (TII)	Total (TE)
Ground points	a	b	$b/(a+b)$	$c/(c+d)$	$(b+c)/(a+b+c+d)$
Non-ground points	c	d			

Table 3
Filter error statistics of fifteen samples.

Sample number	TI	TII	TE
Sample 11	28.99	29.67	29.28
Sample 12	8.41	18.69	13.42
Sample 21	0.43	11.72	2.93
Sample 22	16.53	18.45	17.13
Sample 23	10.66	19.20	14.70
Sample 24	10.49	14.92	11.71
Sample 31	9.85	13.87	11.70
Sample 41	22.67	9.77	16.21
Sample 42	2.89	3.80	3.53
Sample 51	13.18	9.24	12.32
Sample 52	26.01	29.25	26.35
Sample 53	15.40	24.91	15.78
Sample 54	4.17	11.83	8.28
Sample 61	5.56	15.84	5.91
Sample 71	15.26	10.17	14.68

There are all ground object points in the grid, and the actual height difference between the points is greater than epsilon 1, so the lower bridge points are classified into the initial terrain category, and only some higher bridge points are filtered out correctly. The height difference between the bridge point classified incorrectly into the initial terrain category and the adjacent real ground point within range 1 is less than range 2, so it is not selected as a mutation point. In the second clustering, the distance between this part of the bridge points and the new class is greater than epsilon 2; therefore, the bridge points with no category after clustering are considered ground points. The above reasons together cause Type II errors.

The above situation also occurs in Sample 42 of the city sample with high-frequency topographic relief. The filtering error of this sample is mainly a Type II error, as shown by the red dot in Fig. 11. Fig. 12 is a shadow relief map generated by unfiltered data, and Fig. 13 is a shadow relief map generated by filtered data. Take the grid in Fig. 14 as an example. In the final

Table 4
Comparison of total filtering error with other filtering algorithms.

Sample	Elmqvist	Sohn	Axelsson	Pfeifer	Brovelli	Roggero	Wack	Sithole	Mean	Deng
S11	22.40	20.49	10.76	17.35	36.96	20.80	24.02	23.25	22.00	29.28
S12	8.18	8.39	3.25	4.50	16.28	6.61	6.61	10.21	8.00	13.42
S21	8.53	8.80	4.25	2.57	9.30	9.84	4.55	7.76	6.95	2.93
S22	8.93	7.54	3.63	6.71	22.28	23.78	7.51	20.86	12.66	17.13
S23	12.28	9.84	4.00	8.22	27.80	23.20	10.97	22.71	14.88	14.70
S24	13.83	13.33	4.42	8.64	36.06	23.25	11.53	25.28	17.04	11.71
S31	5.34	6.39	4.78	1.80	12.92	2.14	2.21	3.15	4.84	11.70
S41	8.76	11.27	13.91	10.75	17.03	12.21	9.01	23.67	13.33	16.21
S42	3.68	1.78	1.62	2.64	6.38	4.30	3.54	3.85	3.47	3.53
S51	23.31	9.31	2.72	3.71	22.81	3.01	11.45	7.02	10.42	12.32
S52	57.95	12.04	3.07	19.64	45.56	9.78	23.83	27.53	24.93	26.35
S53	48.45	20.19	8.91	12.60	52.81	17.29	27.24	37.07	28.07	15.78
S54	21.26	5.68	3.23	5.47	23.89	4.96	7.63	6.33	9.81	8.28
S61	35.87	2.99	2.08	6.91	21.68	18.99	13.47	21.63	15.45	5.91
S71	34.22	2.20	1.63	8.85	34.98	5.11	16.97	21.83	15.72	14.68

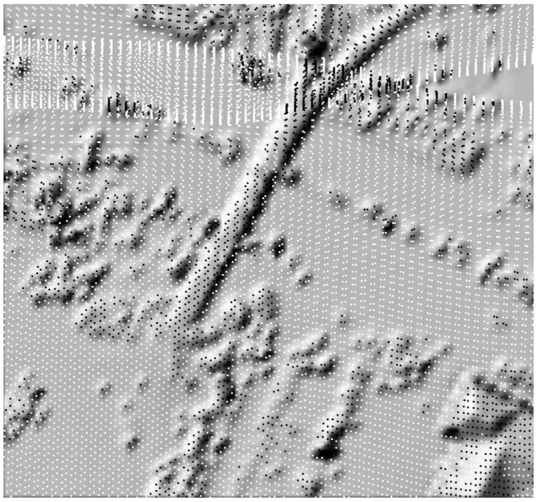

Fig. 3. S21 filtering result.

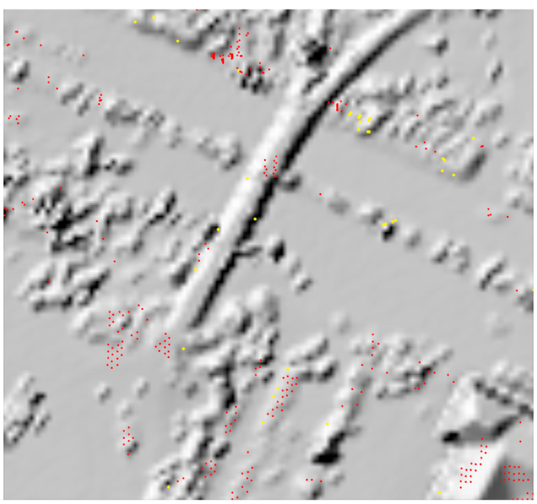

Fig. 4. S21 error result.

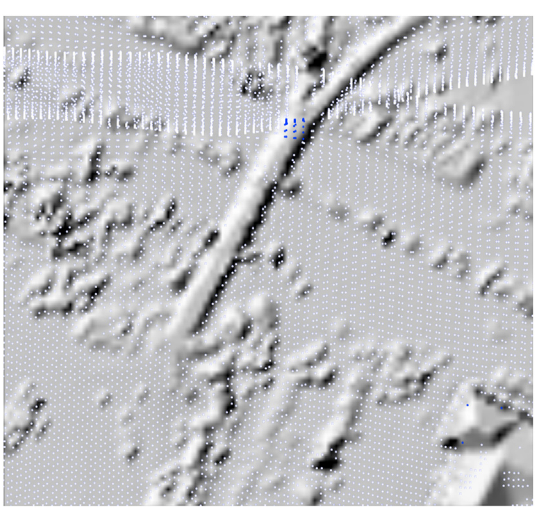

Fig. 5. S21 re-division result.

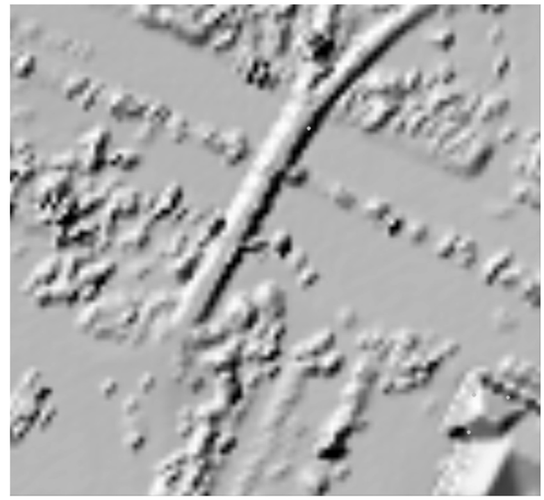

Fig. 6. S21 mutation point.

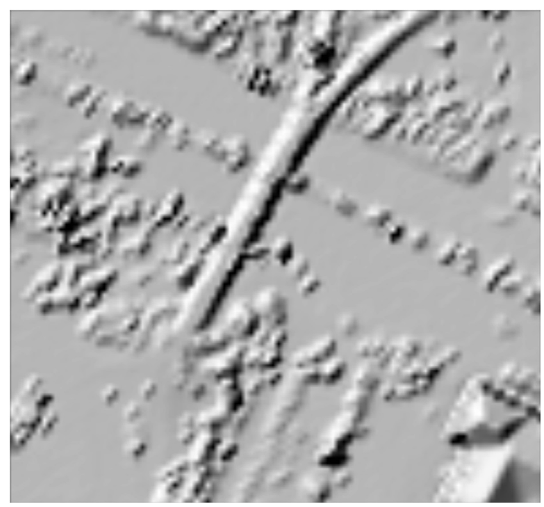

Fig. 7. S21 relief image before filtering.

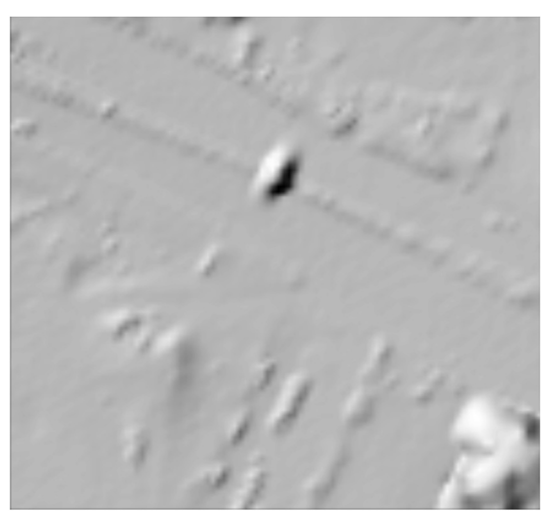

Fig. 8. S21 relief image after filtering.

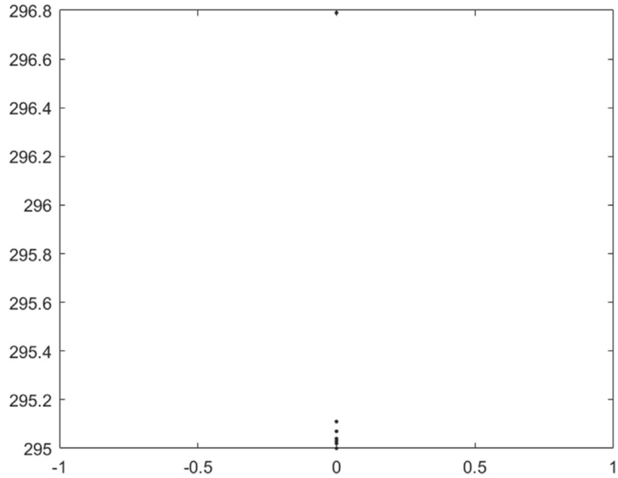

Fig. 9. Sample point elevation distribution map.

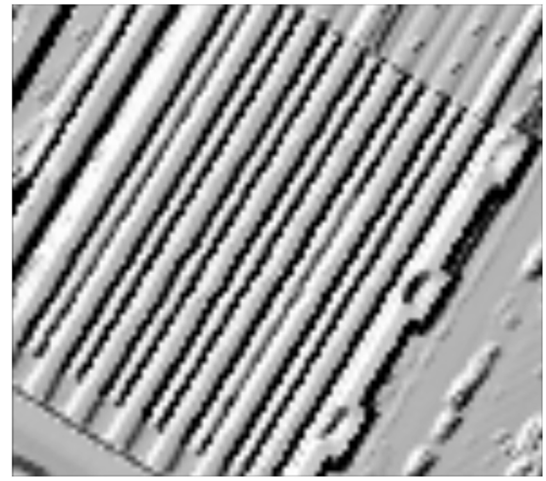

Fig. 12. S42 relief image before filtering.

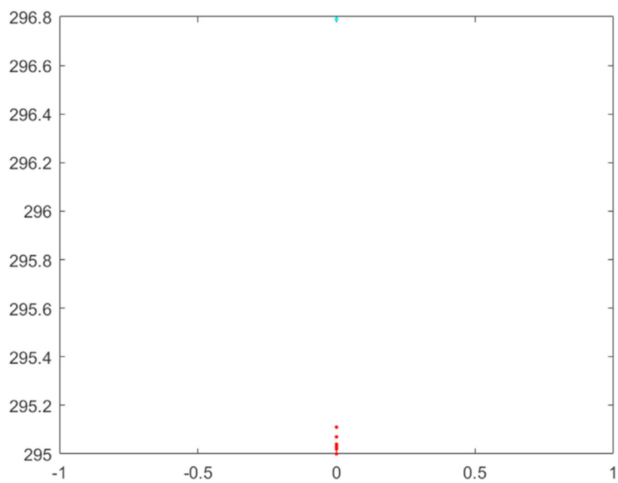

Fig. 10. Sample point classification map.

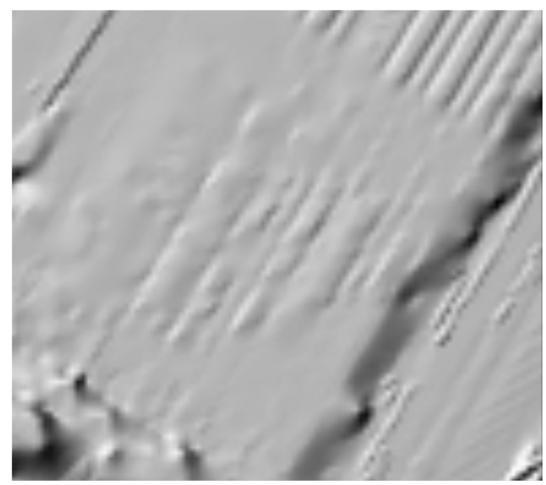

Fig. 13. S42 relief image after filtering.

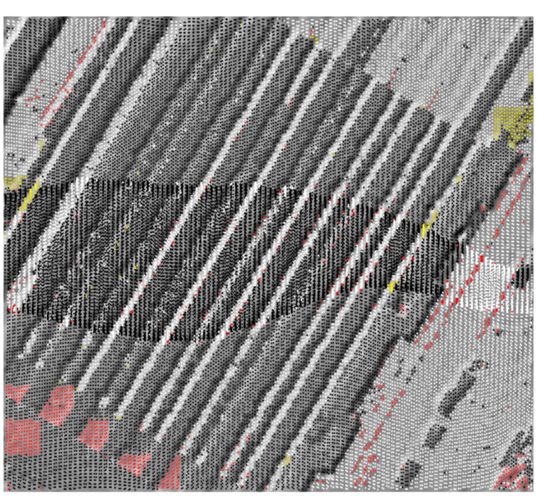

Fig. 11. S42 filtering result.

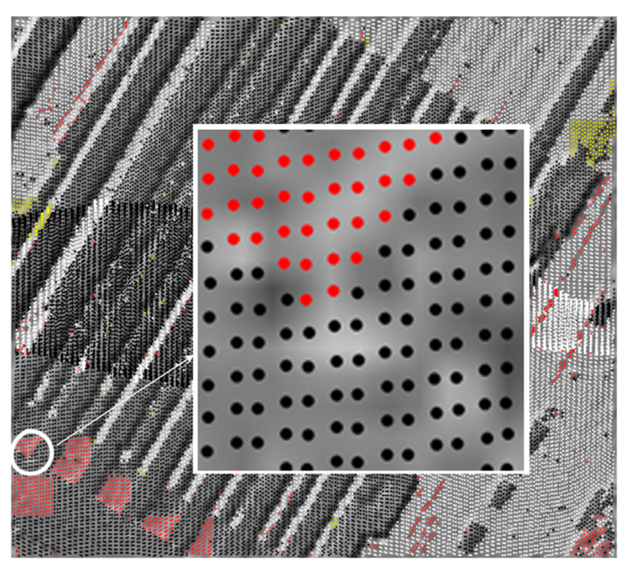

Fig. 14. Example point.

classification result of the grid, there are some Type II error points. Fig. 15 is the elevation distribution map of the points in the grid. Because epsilon 1 is smaller than the actual height difference of the point cloud in the grid, after splitting the points that should belong to the same category, the lower category is mistakenly classified as ground points. Fig. 16 shows the classification results of points in the grid.

4.2.2. Experimental sample 51

Sample 51 is a slope with vegetation from the rural area. The filtering results are shown in Figs. 17 and 18, and the re-division result is shown in Fig. 19. The color meaning of the points in each figure is the same as in the previous section. The orange dots represent the reference ground object points. As shown in Fig. 20, the mutation points indicated by the green dots are all reference ground object points, and most of them are located at the boundary of the ground objects, which shows that most of the mutation points are selected correctly. Fig. 21 is a shadow relief map generated by unfiltered data, and Fig. 22 is a shadow relief map generated by filtered data.

Most of the filtering errors in Sample 51 are caused by Type I errors. The ground points around the vegetation are misjudged as vegetation points. The main reason is that epsilon 1 is small, which causes the ground points that should be classified into the same category to be divided into multiple categories, so only a small part of the ground points are classified into the ground category. The secondary clustering is only performed in the initial terrain class, so the ground points that are incorrectly divided in the first clustering are corrected. The explanation of this process is shown in Fig. 23. If epsilon1 is smaller than the height difference range between the real ground and the ground object, a classification error will occur.

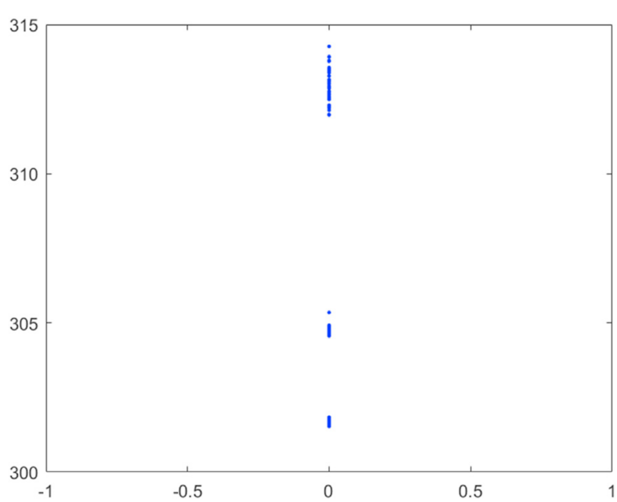

Fig. 15. Elevation distribution map.

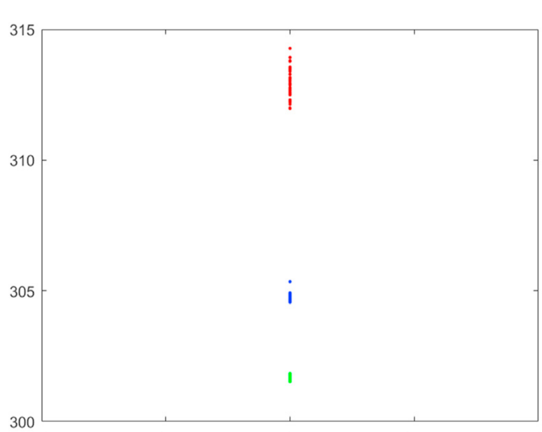

Fig. 16. Classification diagram.

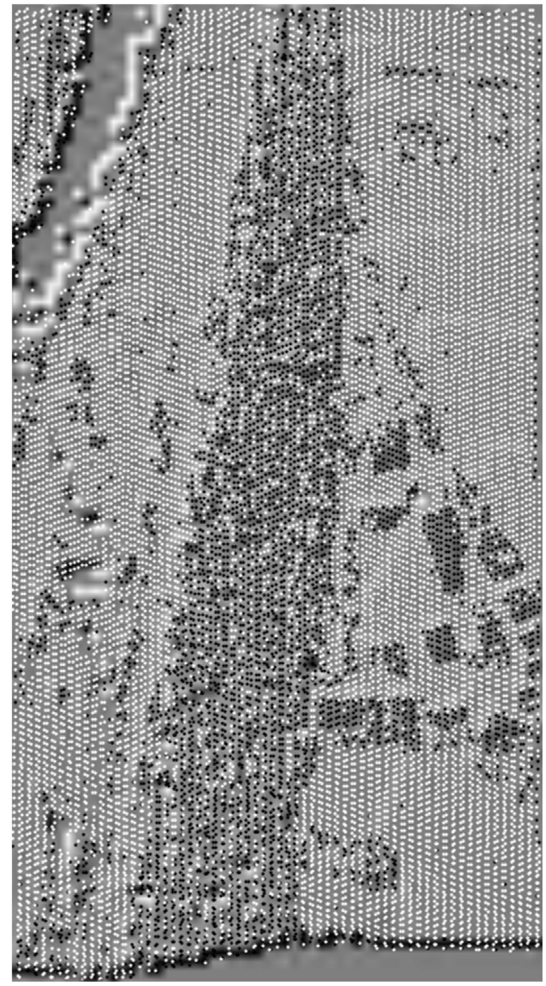

Fig. 17. S51 filtering result.

4.2.3. Experimental sample 53

Sample 53 is a rural sample with discontinuous terrain features. The meaning of each color point in the figure below is the same as that in the previous section. Fig. 24 shows the filtering result of the proposed algorithm. The mutation points represented by the green dots in Fig. 25 show ground points that are misclassified mostly as ground object points, indicating that the selection of most mutation points is wrong. Fig. 26 shows a shadow relief map generated by unfiltered data, and Fig. 27 shows a shadow relief map generated by filtered data. The overall terrain is retained relatively intact, but the changes in ground objects are limited.

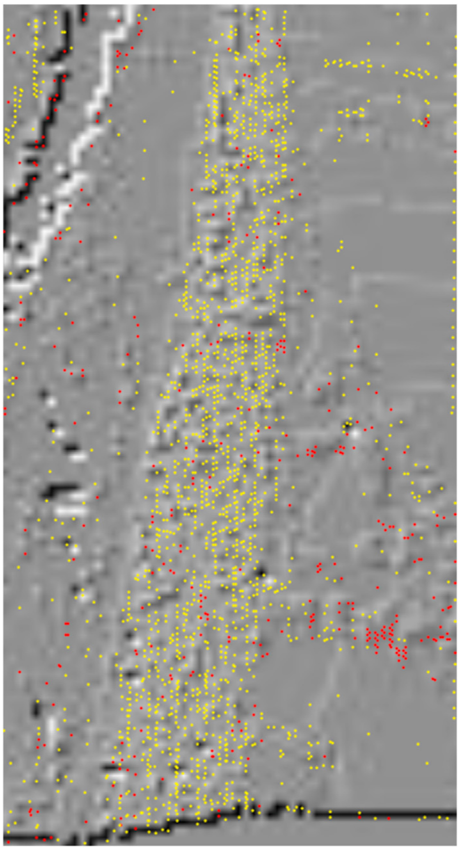

Fig. 18. S51 error result.

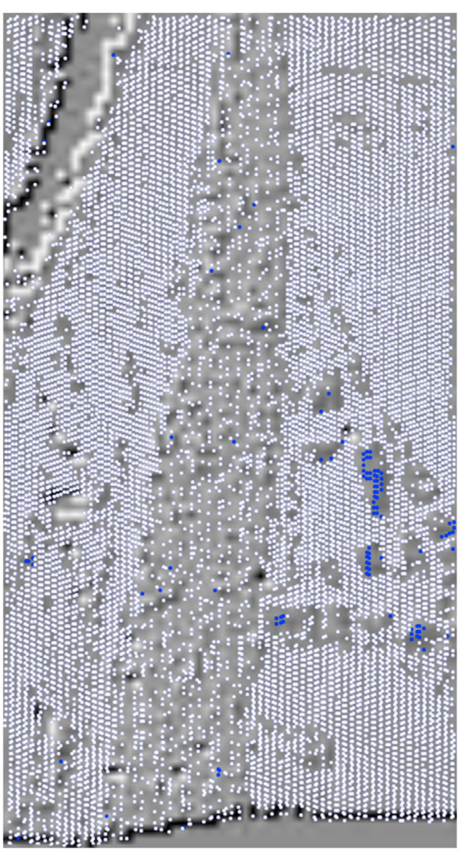

Fig. 19. S51 re-division result.

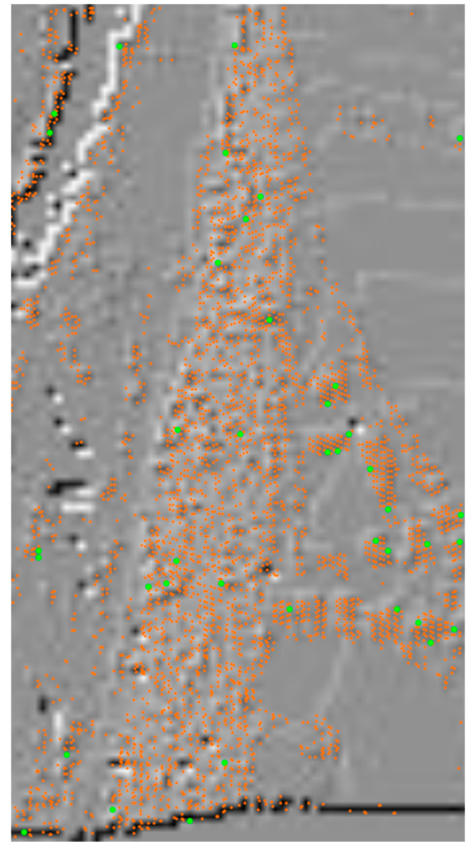

Fig. 20. S51 mutation point.

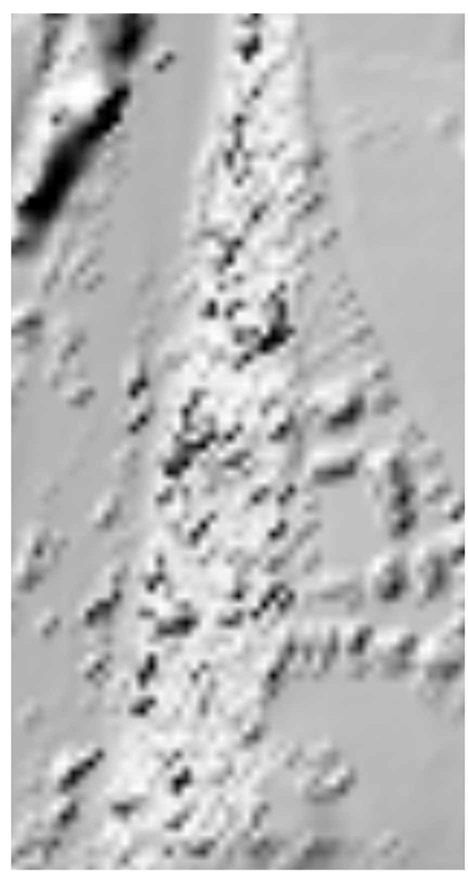

Fig. 21. S51 relief image before filtering.

Fig. 22. S51 relief image after filtering.

Fig. 24 shows that there are many Type I error points, mainly due to the discontinuity of terrain. The height difference between the lower and upper layers is greater than epsilon 1, so the ground points at the edge of the upper layer are misjudged as ground object points. It is also because this discontinuous feature is very similar to the feature of the ground object, and both have height jump within a short distance. Hence, the selection of the mutation point produces is also wrong. Since the initial clustering is done in each small grids, the grids do not interfere with each other, so most of the discontinuous terrains except for the edge part are classified correctly, and the features are retained.

5. Conclusions

This article proposed a novel method of fast classification filtering based on DBSCAN density clustering. After finding the appropriate parameter range according to the actual terrain, the filtering process does not require human intervention. The proposed algorithm is validated with 15 sets of samples, and compared with the other eight classical filtering algorithms; the conclusions are remarked as follows: (1) The proposed algorithm can obtain good filtering results in both rural and urban areas. It can remove vegetation coverage, bridges, and buildings, it can also retain topographic relief in discontinuous terrain areas very well. (2) The proposed algorithm relies on the "density connection" of the ground point clouds. When the "density connection" between the point clouds is satisfied under the set threshold, even if the elevation difference between the point clouds is large and the slope change is large too, the high point clouds can be recognized as ground points. If the threshold is set too large, the object points may be "density connected" with the ground points, resulting in the classification of the object points as ground points. If the threshold is too small, the ground points may be classified as object points. (3) Through reclassification, the object points that were first classified as ground points can be eliminated. However, since the initial clustering center of the reclassification is selected by using the sudden change of the object points in space, this will cause the ground points on the edge of the discontinuous terrain to be mistaken for the object points. After the secondary clustering, they are eliminated as the object

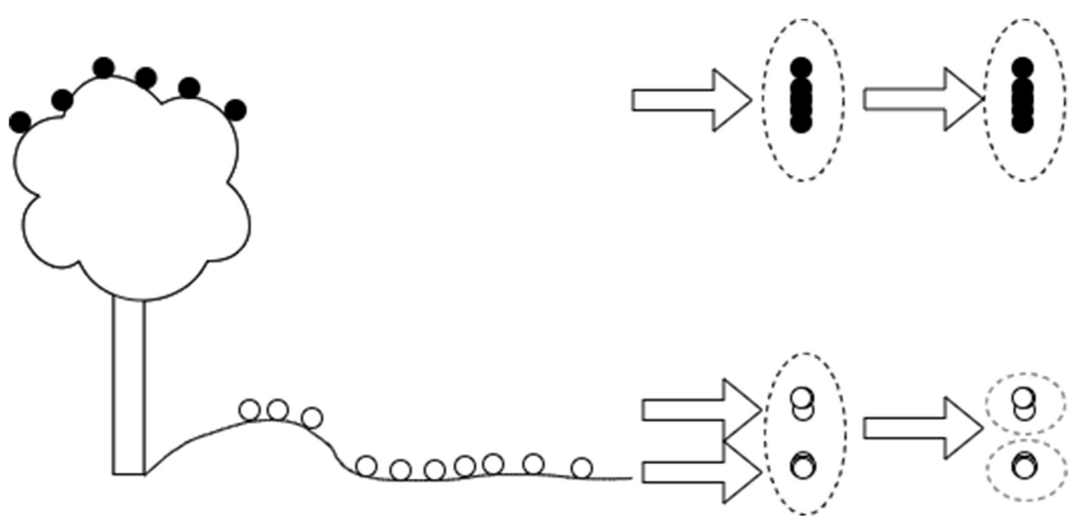

Fig. 23. Classification error diagram.

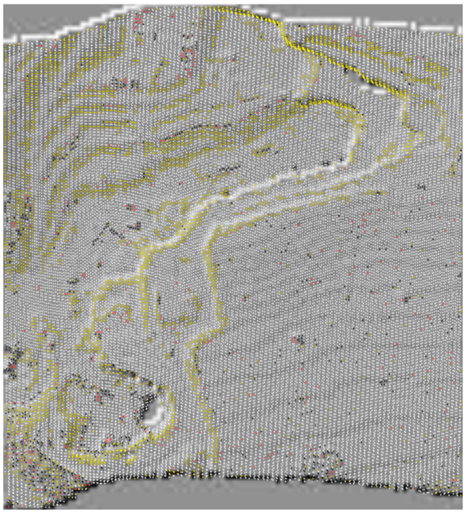

Fig. 24. S53 filtering result.

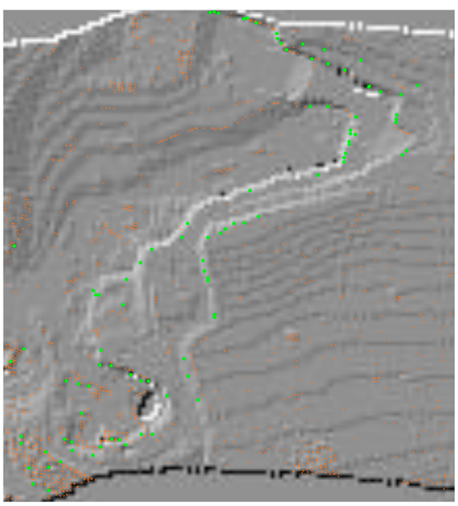

Fig. 25. S53 mutation point.

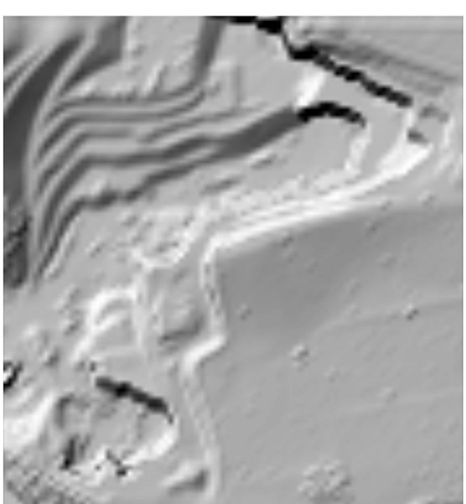

Fig. 26. S53 relief image before filtering.

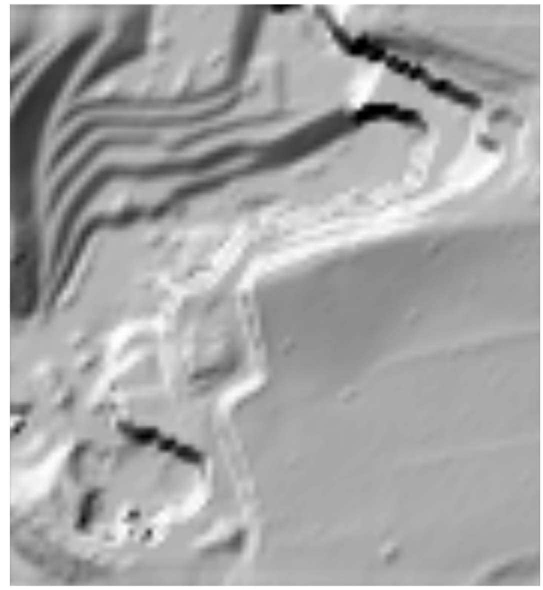

Fig. 27. S53 relief image after filtering.

points. So the proposed algorithm still has filtering limitations at the edges of discontinuous terrain.

Foundation support

The Natural Science Foundation of Hunan Province, China (No. 2020JJ4601); Open Fund of the Key Laboratory of Highway Engineering of Ministry of Education (No. kfj190203).

Conflicts of interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled “A novel Fast Classification Filtering Algorithm for LiDAR Point Clouds Based on small Grid Density Clustering”.

References

[1] L. Sui, B. Zhang, Principle and trend of airborne laser scanning remote sensing, Journal of Geomatics Science and Technology 2 (2006) 127–129.
[2] B. Yang, F. Liang, R. Huang, Progress, challenges and perspectives of 3D LiDAR point cloud processing, Journal of Geodesy and Geoinformation Science 46 (10) (2017) 1509–1516.
[3] S. Chen, X. Cheng, The principle and application of airborne LIDAR, Engineering of Surveying and Mapping 1 (2007) 27–31.
[4] X. Huang, H. Li, X. Wang, Filter algorithms of airborne LiDAR data: review and prospects, Journal of Geodesy and Geoinformation Science 38 (5) (2009) 466–469.
[5] G. Sithole, G. Vosselman, Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds, ISPRS J. Photogrammetry Remote Sens. 59 (1/2) (2004) 85–101.
[6] G. Vosselman, Slope based filtering of laser altimetry data, International Archives of the Photogrammetry and Remote Sensing 33 (Part B3) (2000) 935~942.
[7] G. Sithole, Filtering of laser altimetry data using a slope adaptive filter, International Archives of the Photogrammetry and Remote Sensing 34 (Part3/W4) (2001) 203~210.
[8] Shaopeng Ding, Rufei Liu, Yongning Cai, Peng Wang, A point cloud adaptive slope filtering method considering terrain, Remote Sens. Inf. 34 (4) (2019) 108–113.

[9] R. Wack, A. Wimmer, Digital terrain models from airborne laser scanner data – a grid based approach, International Archives of Photogrammetry and Remote Sensing 34 (3) (2002) 293~296.

[10] Qiguang Miao, Xue Guo, Jianfeng Song, et al., LiDAR point cloud data with morphological filter algorithm based on region prediction, Laser & Opto-electronics Progress 52 (1) (2015) 67–73.

[11] M. Elmqvist, Ground Estimation of Laser Radar Data Using Active Shape Models: OEEPE Workshop on Airborne Laser scanning and Interferometric SAR for Detailed Digital Elevation Models, Royal Institute of Technology Department of Geodesy and Photogrammetry, Stockholm, 2001.

[12] X. Liu, Y. Chen, C. Liang, et al., Airborne laser scanning point clouds filtering method based on the construction of virtual ground seed points, J. Appl. Remote Sens. 11 (1) (2017), 016032.

[13] F. Li, A Research on Filtering and Classification of Airborne LiDAR Point Clouds, China University of Mining and Technology, Beijing, 2013.

[14] G. Sithole, G. Vosselman, Filtering of Airborne Laser Scanner Data Based on Segmented Point Clouds. Laser Scanning 2005, Enschede, Netherlands, 2005.

[15] N. Chehata, N. David, F. Bretar, LIDAR data classification using hierarchical K-means clustering.The international archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37 (Part B3b) (2008) 325–330.

[16] M. Ester, H.-P. Kriegel, J. Sander, et al., A Density-based algorithm for discovering clusters in large spatial databases with noise. Portland, in: Proceedings of the 2nd ACM SIGKDD, 1996, pp. 226–231.

[17] W. Qian, X. Gong, A. Zhou, Clustering in very large databases based on distance and density, J. Comput. Sci. Technol. 18 (1) (2003).

[18] S. Zhou, A. Zhou, J. Cao, et al., A fast density-based clustering algorithm, J. Comput. Res. Dev. 11 (2000) 1287–1292.

[19] Q. Rong, J. Yan, G. Guo, Research and implementation of clustering algorithm based on DBSCAN, Comput. Appl. 4 (2004) 45–47.

[20] J. Sun, J. Liu, L. Zhao, Clustering algorithms research, J. Software 1 (2008) 48–61.

[21] X. Ma, Y. Tang, Hierarchical clustering algorithm research, Computer Science 34 (7) (2008) 34–36, 2008.

[22] A.L.N. Fred, J.M.N. Leitão, Partitional vs hierarchical clustering using a minimum grammar complexity approach, Lect. Notes Comput. Sci. 1876 (2000) 193–202.

[23] Z. Huang, A fast clustering algorithm to cluster very large categorical data sets in data mining, in: Proc. of the SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery. Tucson, 1997, pp. 146–151. http://www.informatik.uni-trier.de/~ley/db/conf/sigmod/sigmod97.html.

[24] Z. Huang, Extensions to the k-means algorithm for clustering large data sets with categorical values, Data Min. Knowl. Discov. 2 (3) (1998).

[25] Eden W.M. Ma, Tommy W.S. Chow, A new shifting grid clustering algorithm, Pattern Recogn. 37 (3) (2003).

[26] A.H. Pilevar, M. Sukumar, GCHL: a grid-clustering algorithm for high-dimensional very large spatial data bases, Pattern. Recogn. Lett. 26 (7) (2004).

[27] Mirco Nanni, Dino Pedreschi, Time-focused clustering of trajectories of moving objects, J. Intell. Inf. Syst. 27 (3) (2006).

[28] Derya Birant, Kut Alp, ST-DBSCAN: an algorithm for clustering spatial–temporal data, Data. Knowl. Eng. 60 (1) (2006).

[29] C. Tsai, H. Wu, C. Tsai, A new data clustering approach for data mining in large databases, in: International Symposium Parallel Architectures, Algorithms and Network, Manila, the Philippines, 2002, pp. 278–283.

[30] G. Wang, G. Lin, Improved adaptive parameter DBSCAN clustering algorithm, Comput. Eng. Appl. J 56 (14) (2020) 45–51.

[31] W. Li, S. Yan, Y. Jiang, Research on method of self-adaptive determination of DBSCAN algorithm parameters, Comput. Eng. Appl. J 55 (5) (2019) 1–7+148.

[32] J. Gan, Y. Tao, DBSCAN revisited: Mis-claim, un-fixability, and approximation, in: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, ACM, 2015, pp. 519–530, 2015.

[33] A. Lulli, M. Dell'Amico, P. Michiardi, et al., NG-DBSCAN: scalable density-based clustering for arbitrary data, Proceedings of the VLDB Endowment 10 (3) (2016) 157–168.

[34] Y. Chen, W. Ruys, G. Biros, KNN-DBSCAN: a DBSCAN in high dimensions, 2020. https://arxiv.org/abs/2009.04552.

[35] Y. Chen, L. Zhou, S. Pei, et al., KNN-BLOCK DBSCAN: fast clustering for large-scale data, IEEE Transactions on Systems, Man, and Cybernetics: Systems 51 (2021) 3939–3953.

[36] Y. Chen, L. Zhou, N. Bouguila, et al., BLOCK-DBSCAN: fast clustering for large scale data, Pattern. Recogn. 109 (2020) 107624.

[37] Sithole George, Vosselman George, Report: ISPRS comparison of filters, 2003. http://www.itc.nl/isprswgIII-3/filtertest/.

[38] X. Zhao, Q. Guo, Y. Su, Improved progressive TIN densification filtering algorithm for airborne LiDAR data in forested areas, ISPRS J. Photogrammetry. Remote. Sens. 117 (2016).

[39] X. Zhu, C. Wang, X. Xi, Hierarchical threshold adaptive for point cloud filter algorithm of moving surface fitting, J Geodesy Geoinf Sci 47 (2) (2018) 153–160.

[40] M. Zhao, J. He, J. Tian, LiDAR data filtering method based on improved progressive multi-scale mathematic morphology, Acta. Opt. Sin. 33 (3) (2013) 292–301.

Xingsheng Deng received the Ph.D. degree from Wuhan University, Wuhan, China, in 2007. He is currently an associate professor with Changsha University of Science & Technology, Changsha, China. His current research focuses mainly involve Photogrammetry and Remote Sensing data processing. E-mail: whudxs@163.com