Acta Technica 62, No. 3B/2017, 1–16

Research on-board LIDAR point cloud data pretreatment

Peng Cang[1], Zhenglin Yu[1], Bo Yu[2,3]

Abstract. In view of the on-board LIDAR system problems existing in point cloud data pretreatment process, analyzes of the on-board LIDAR system to obtain the spatial distribution characteristics of point cloud data, based on the gradient difference of on-board LIDAR point cloud data filtering method, and analyzes the concrete process of its implementation, at the same time, this paper expounds the current several mature point cloud facade segmentation method, and mainly discusses the vehicle based on random sampling algorithm of LIDAR point cloud building facade segmentation is implemented. At the end of the article with the introduction of point cloud data comparative analysis of the experiment, the actual block further demonstrates the effectiveness of the method given in this paper.

Key words. On-board LIDAR, point cloud data filtering, grade difference, point cloud facade segmentation, random sampling algorithm.

1. Introduction

Data preprocessing on-board LIDAR system mainly includes on-board LIDAR point cloud data filtering and building facade segmentation [1]-[3]. When On-board LIDAR scanning on city streets and access to the great amount of point cloud data, the original point cloud data contains the building point cloud and includes the streets, feature point cloud (trees, street lamp, etc). This will affect buildings and other accurate features because of blind point cloud data on-board LIDAR and non-measured analyte. Therefore, this will affect buildings and other features of accurate classification. At the same time, it also make the point cloud from a large number of point cloud data to extract building information more difficult [4], [5]. Features based on on-board LIDAR point cloud data, this paper proposes a point cloud data filtering method based on gradient difference, at the same time, by the consistency of

[1]School of Mechatronic Engineering, Changchun University of Science and Technology, Changchun, 130000, China

[2]School of Mechatronic Engineering, Changchun Institule of Technology, Changchun, 130000, China

[3]Corresponding author; e-mail: yubo745@163.com

the random sampling algorithm, by setting the corresponding parameter threshold, the building of filtering facade segmentation of point cloud.

2. The point cloud space distribution characteristics

The filter can extract the target building facade from laser scanning data point cloud, and eliminate the lines of trees, pedestrians, vehicles and other obstructions. And laser can go through the transparent object after the formation of the scattered points. The influence of target scanning point in the block can be divided into three classes of ground, building facade and other features. On-board LIDAR point cloud, though on the space distribution of discrete point cloud, but as a result of these different scan point object in the real environment, point cloud in different objects in the same line profile will show different distribution characteristics (Fig. 1).

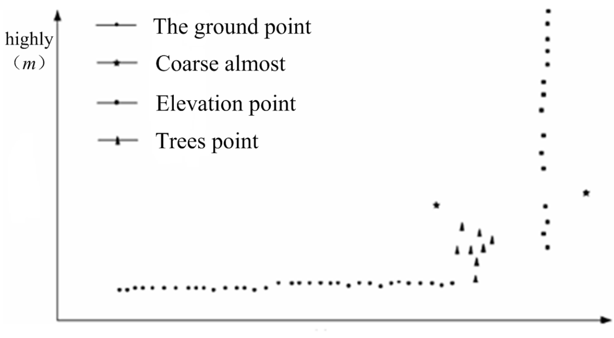

Fig. 1. Distribution of scanning point cloud

(1) The distribution features of the building facade points:

Generally speaking, the basic perpendicular to the ground and building facade at right corner. On the same line, the adjacent image point on the horizontal axis difference is very small, on the vertical axis vertical slope approximation, presents a piecewise smooth distribution of vertical linear approximation.

(2) The distribution characteristics of ground points:

In vehicular laser scanning point cloud, the ground point is generally flat, elevation values are relatively small and little change, the general level of a linear distribution on the scan line.

(3) Other distribution characteristics of feature points:

Other features including lines of trees, electric pole, flower beds and other targets. Due to irregularity of the structure itself, reflected in the lines of trees in discrete state local area on the scan line. Their targets at the same time also are small, reflected in the pole and fewer data points, flower bed so point cloud on these goals is sparse and irregular distribution, but it is also possible that they present a linear distribution on a small scale.

3. Based on gradient difference of on-board LIDAR point cloud data filtering method

Because of the object can be measured on-board LIDAR, its access to the original point cloud data which spatial distribution along the line. However, different methods are used in scanning space with different distribution characteristics. Because the article is aimed at modeling, digital city and get point cloud data is true city block, including: point cloud data, building of point cloud data, street trees and other feature point cloud data point cloud data. In general, perpendicular to the ground to build buildings and its edge horn is right Angle. On the same line, get the point cloud data of adjacent points in a horizontal direction difference is very small, in the vertical direction of the slope is close to 90°, embodied in the continuous piecewise smooth approximation vertical linear distribution. Point cloud data is generally flat, street adjacent point elevation value is small, and the elevation difference is close to zero, at the same horizontal arrangement on line. For trees and other ground objects (lamps, vehicles, etc.), due to its structure is not unity, regularity, point cloud data in the same line of scans in discrete status display. Bicycle, street lamps and other features due to its own volume is too small, get the point cloud data is small, the feature of point cloud data density is small and irregular distribution, but may also be a linear distribution in small scope.

Based on the above analysis, although the LIDAR point cloud data can be collected automotively, the overall distribution in three-dimensional space is discrete. However, in the process of actual scanning, because scanning feature type is different, its get point cloud data in the same show different features within the line profile distribution. Based on the characteristics of on-board LIDAR point cloud data, a slope difference of on-board LIDAR point cloud data filtering method is proposed. First, the noise in the initial point cloud data (discrete points) can be removed by using the gradient difference which scans lines into different small line collection, and the discrete point cloud data can be removed in the small line segments. After removing noise from small line segments, the attributes can be combined with point cloud data classification. Finally, using local linear regression method, to improve the classification results, improve the classification accuracy, so as to get high quality point cloud data filtering results.

3.1. The point cloud in addition to dry

On-board LIDAR system scans the object which is measured in accordance with the line scan mode of access to the original point cloud data. In the scanning process in accordance with the analyte per column on a line by line scanning, it is due to random errors. So to get the original points existing in point cloud data noise, the noise points seriously affect the accuracy of modeling. Therefore, the noise points must be removed. Usually, noise point elevation value is higher, on the same line, the elevation difference between the noise points adjacent points than other adjacent point of elevation difference value of ground objects. In accordance with the above features, at some point can be calculated with the elevation difference between the adjacent point around the average method to determine whether the point for the noise points. With P set as the threshold value, if some points of the elevation difference among the adjacent, the mean value is greater than the threshold value of P then the point can be seemed as noise point. The calculating formula of threshold P is

$$P = D + 3\lambda\,. \tag{1}$$

where λ is an error of the laser angle, and D is neutral point cloud data..

3.2. The point cloud segmentation based on gradient difference

Analysis of automotive original LIDAR point cloud data of the spatial distribution law of point cloud data is usually not that street there are mutations and terrain elevation value mutation, street point cloud data can be seen as smooth linear approximation; And a point cloud data structure usually is perpendicular to the horizontal plane distribution, building point cloud data can be seen as approximately perpendicular to the horizontal plane smooth and straight. Therefore, the spatial distribution characteristics of buildings and streets can be seemed as the basis of a processing line, where the line can be divided into different small line segments. These different small line segment set contains the point cloud point clouds and streets, buildings are not small line segments within a collection of point cloud data judgment for discrete point cloud data, to determine the discrete point cloud other feature point cloud.

Point cloud segmentation based on gradient difference, in accordance with the on-board LIDAR scanning way, calculating the geometric relationship between adjacent laser spot which are point set, scan lines into smaller segments of different small line feature point set contain different attributes, thereby removing discrete point cloud. The above method is the key to using the gradient difference to determine line profile distribution of adjacent points in the relationship, so as to accurately screening. In the same line profile, consisting of two small line gradient value computation formula is

$$\mu_n = \arctan\left[\frac{Z_n - Z_{n-1}}{\sqrt{(X_n - X_{n-1})(X_n - X_{n-1}) + (Y_n - Y_{n-1})(Y_n - Y_{n-1})}}\right]. \tag{2}$$

Here, $-\frac{\pi}{2} < \mu < \frac{\pi}{2}$, X_n, Y_n, Z_n and X_{n-1}, Y_{n-1}, Z_{n-1} denote the space three-dimensional coordinates of adjacent points. By the formula (2), the grade difference formula is

$$\Delta\mu_n = |\mu_{n+1} - \mu_n|\,. \tag{3}$$

The slope difference definition are as is shown in Fig. 2. Symbols P_1, P_2, P_3 are adjacent points, α, β are the slope angles.

In fact, the linear distribution of point cloud data of slope difference between each point is equal to 0. In fact, the small straight lines are not standard, so it is necessary to set threshold which approximate to synthetic small segments linear.

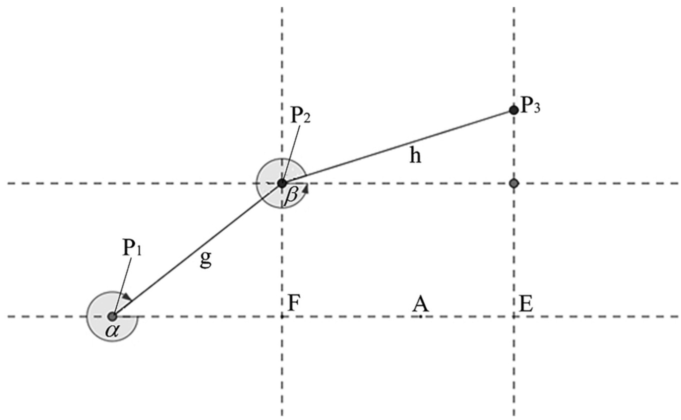

Fig. 2. The gradient difference definition figure

But, because of the trees, street lamp is linear distribution features, such as a small point cloud also because some small line segments within a collection of points is not buildings and streets. In order to avoid the occurrence of this kind of error, set a threshold limit the number of each small line within a collection point, determine the set points is greater than the small line segments to buildings, streets, points, points less than the small line segment set as discrete points. In a word, two thresholds and restrictions can be set by using segmentation algorithm, the set threshold values are as follows:

(1) Considering the point cloud point clouds and street building elevation values changed little, elevation values between adjacent points close, so that the slope of the buildings and streets should be between 0° and 15°. Set a threshold value for 5° to 15°, synthesize small line approximation fitting straight line segments.

(2) Due to different scanning city blocks environment, reference threshold setting is different, so, according to the actual scan density and building profile feature information to determine the threshold. Threshold is too easy to cause the lack of building profile details, the threshold is too small, easy to cause the misjudgment of discrete points. Under normal circumstances, the threshold should be slightly smaller than in detail on the side of the buildings within the same scan line profile (window, balcony, etc.) the point cloud of points. The point cloud segmentation algorithm based on gradient difference process is shown in Fig. 3.

3.3. The point set classification processing small segments

After using the above algorithm, the vehicular original LIDAR point cloud data classification in the early has been completed. Classification after small segment point collection point cloud contained within buildings, streets point cloud, and as a result of misjudgment of small amounts of other feature point cloud (with a small number of continuous distribution points cloud features). In order to assure the precision of classification, the need for classification of small line point set is

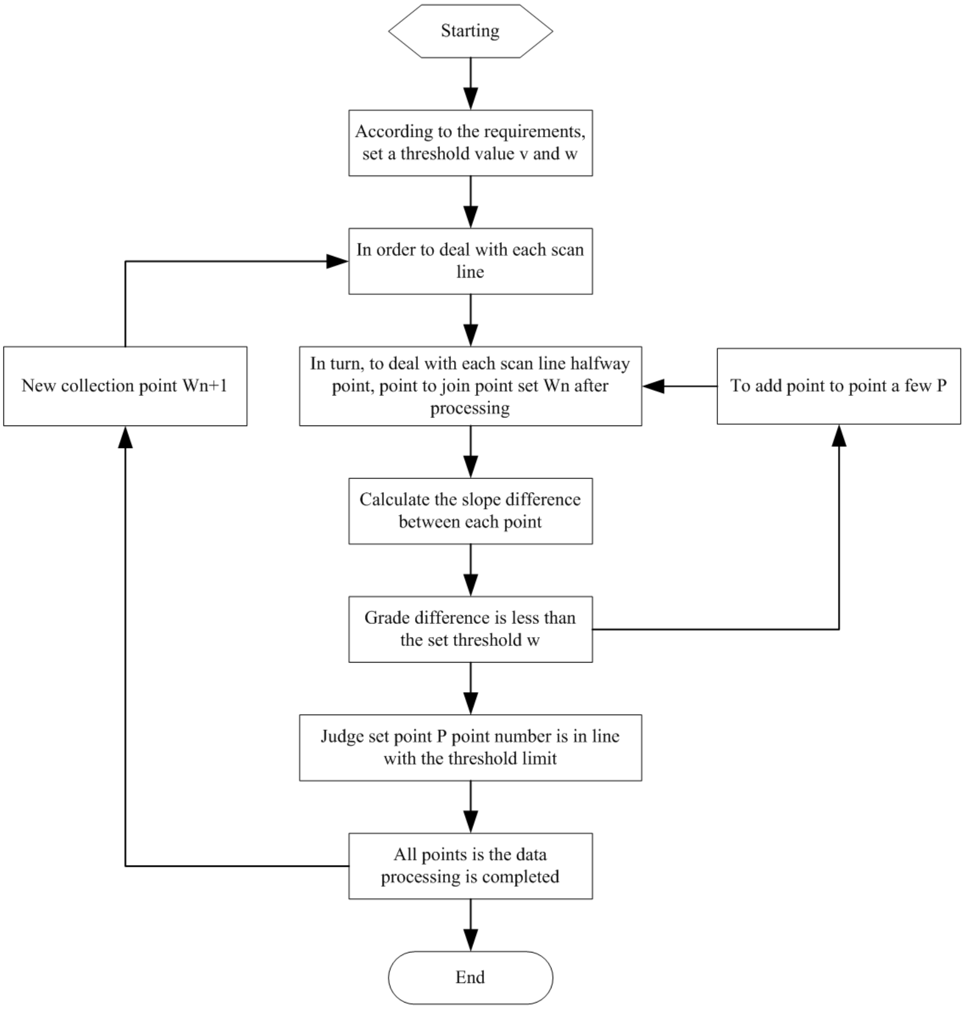

Fig. 3. The segmentation algorithm based on gradient difference process

classified so as to obtain building small segment point cloud data collection, street small line segment set point cloud data, other small feature point cloud data which including line slope and small line coordinates. Small segment slope through a small point within the line fitting, line coordinates for small line segments within the first point coordinates of the average. A small segment slope to K_n where n is a number of line segments within a collection points. Set the first point coordinates of X_1, Y_1, Z_1, end point coordinates of X_n, Y_n, Z_n, the small line coordinates X_n'', Y_n'', Z_n'', the computation formula is

$$(X_n", Y_n", Z_n") = \left(\frac{X_1 + X_n}{2}, \frac{Y_1 + Y_n}{2}, \frac{Z_1 + Z_n}{2} \right). \tag{4}$$

According to the characteristics of the small line point set by setting threshold value method is proposed for small line segment set points for precise classification. Because of on-board LIDAR scanning, building facade scanning center distance is far, so, get point cloud data of building also far distances from the center of the scan lines, and basic moderate city streets, tilt Angle is within 10 °, so street elevation value of the point cloud data is low. According to the characteristics of threshold setting elevation H, respectively, and distance threshold S, if

$$Z_n" - \min(Z") \leq H\,. \tag{5}$$

The point cloud data which meets the formula (5) can be judged as street.

$$|Y_n" - \max(Y")| \leq S\,. \tag{6}$$

Assumptions that satisfied formula (6) point cloud can be judged as building. Assumptions, and urban street width of 10 m, street gradient within the 10 °, street after calculated elevation value between 0 and 60 cm, so the distance threshold S take 60 cm. Since city buildings more side a balcony, the window, such as terrain, the prominent features of projection distance, generally between 0 and 1.5 m, to ensure that the modeling precision to avoid information loss caused by classification structure detail, S distance threshold set to 1.5 m.

3.4. Correction

After dealing with the small segment point set classification, has achieved the original point cloud point cloud point cloud data classification for building, street, other feature point cloud. In other feature point cloud data, however, it may still contains a small amount of point cloud building. Because in a small segment point set classification processing phase, it will be part of the building wrong point cloud into other feature point cloud. In order to ensure the accuracy of the late building three-dimensional reconstruction model, need to get as much as the building of point cloud data. So, the small segment classification of results is corrected.

On-board LIDAR to get point cloud of building much for building facade point cloud, part of the structure is complex and has multiple layers, inside the facade of the all profile point cloud data obtained by combined cannot complete the processing results of fixed, so choose one dimensional linear equation for all small line segments in the collection points in each section separately linear fitting. Is within a specified size profile window, that facade meets the following equation

$$z_{ij} = a_i + b_i y_{ij}\,. \tag{7}$$

In the formula, $i = (1, 2, ..., n)$, n being the number of small line segments in the section point set, j represents for the small line points within a collection of serial number, a_i and b_i mean small line point within a collection point to synthetic straight after the resulting coefficient. Other use formula (7) obtain profile feature points after fitting elevation, combined with the calculated height of the original point cloud fitting residual error, if the residual error is less than three times the standard deviation of the elevation, point is judged to be building.

4. On-board LIDAR point cloud data building facade segmentation method

Firstly, the modeling of the building is needed in the last section of the point cloud data filtering classification on the basis of the point cloud data for building facade segmentation, then the building elevation information can be extracted. Segmentation of point cloud data commonly used method is roughly divided into the following kinds:

(1) The region growing method: this method first extracted three points from the point cloud to synthesis of a plane, find the minimum fitting residual error of a flat as initial plane, then using the law of regional growth, will be in near extraction with similar geometric feature points divided into initial plane, so as to obtain the initial plane under the limit of the whole surface area. For the rest of the filtered point data to continue using region growing method, until all the points data screening is complete, you can get rest after segmentation of point cloud data another 1.

(2) Clustering methods: grouping of point cloud data in the feature space, feature space contains some of the normal vector, strength, and the roughness of the local feature vector, due to the points on the same table variable has the same features, the same on the surface of a point in the feature space can form a populated area, so, using the method of clustering can be extracted from point cloud data surface area 2.

(3) Consistency random sampling algorithm: in order to get effective sample data algorithm 3, a mathematical model are used based on a set of samples contain abnormal data sets. Random sampling algorithm assumes that the samples contained in consistency can be described by mathematical model of the data points and cannot be described by mathematical model data points (noise). Usually, these noise points due to the calculation error, scan error and setting error. In order to get effective sample data algorithm 3, a mathematical model are used based on a set of samples contain abnormal data sets. Usually, randomly selected from three points as the initial point in the plane, and calculate the selection point of the plane parameters, set a threshold value of the initial point cloud data filtering, so as to get all the points in the plane, the above steps Nth cycle. Will get results every time compared with the last time the result, if the results better than last time, replace the last parameter. The algorithm includes four types of input parameters: the point on the surface of the peace distance threshold, three-dimensional coordinates of points, the same plane biggest contains points, to capture at least one of the best in the Nth sampling observation of minimum probability. Basic steps as shown in Fig. 4, the algorithm mentioned the confidence probability is to capture at least one of the best in the Nth sampling observation of minimum probability, if the sampling number for N times, data error rate, the confidence probability P_0 computation formula is as follows:

$$P_0 = 1 - (1 - (1 - \eta)^n)^N \,. \tag{8}$$

5. Random sampling algorithm based on segmentation of on-board LIDAR point cloud building facade

Considering the random sampling algorithm good processing ability, even if the original data contains more than fifty percent of the abnormal data, ideal results can be still obtained by using algorithm processing. After classification in this paper, the vehicle LIDAR point cloud data of the facade can be discretized by using a random sampling algorithm. First of all, because most of the structures is simple and the angle is more than 90 ° more, it is necessary to establish a mathematical model of the planar parameters. So the contour line in X axis and Y axis direction in three-dimensional space coordinates perpendicular to each other. In general, it should be satisfied at the same plane point:

$$Ax + By + Cz = D\,. \tag{9}$$

In this formula, x, y, z are interior point of a planar 3D space coordinates; A, B, C are the unit normal vector of surface and D stands for the origin of coordinates.

From the classification of point cloud data to mention point cloud plane parameters of rest is to calculate each another. In the form of a basic matrix of the parameters of the plane, so it will be point cloud segmentation problem is transformed into matrix estimation problem. Assumptions, set at a point on the building facade as $S = (x_n, y_n, z_n$, n is the total number of points in the collection of all, in the form of a matrix is expressed as

$$[x_n, y_n, z_n, -1]\, F = 0\,. \tag{10}$$

The basic matrix F is

$$F = [a, b, c, d]^{\mathrm{T}}\ . \tag{11}$$

To sum up, there are three basic matrix F freedoms, which is used to compute the basic matrix needs at least three points of data. Assumption, randomly selected from all three points for in-plane point and calculate the initial plane parameters, initial plane parameters are selected according to other points in the plane, such not only improves the speed of the data processing, and the biggest possible to eliminate the influence of noise points and plane outside. Parametric mathematical model was set up after the completion of the need to design the corresponding criterion which is based on the Euclidean distance used to determine standards and related design. Instruction in $P = (x, y, z)$, flat $P_0 = (a, b, c, d)$ is the Euclidean distance to point P to the plane of the P_0

$$D(P, P_0) = |ax + by + cz - D|\ . \tag{12}$$

Usually, plane point distance plane is 0, but a collection of plane is actually the point mentioned above is not a real plane, so by setting threshold gamma synthesize the approximate fitting plane. In building facade point cloud data, and the corresponding plane distance if is greater than the set threshold gamma, the decision point for the out-of-plane points, if less than threshold gamma is judged to be

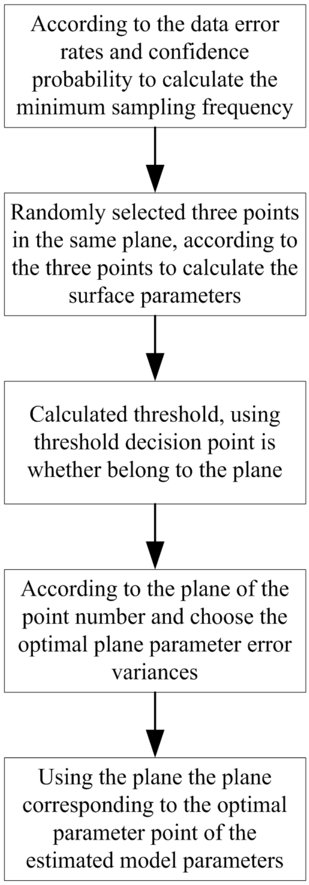

Fig. 4. Random sampling process consistency algorithm

plane. This according to threshold determination method of threshold setting is very strict, if the threshold is too large, easy to build plane into excessive segmentation, the threshold is too small of surface corrosion. By formula that random sampling algorithm of time consistency with the exception of point cloud data contains data related, the smaller the threshold is set to screen out the more data contains abnormal points. In order to guarantee the feasibility of the algorithm, random sampling consistency algorithm need cycles increase rapidly. In setting the threshold value, also need to consider building their own facade shape, for detail information too much of the building facade, the threshold setting to relatively strict, and fewer details of building facade threshold setting is loose. By the judge, can be in conformity with the threshold selection point to plane point, however, only by Euclidean distance does not guarantee segmentation accuracy completely, so the r radius density on misjudgment point in the graphics screen out again. As shown in Fig. 5, the circle of radius is r, it is obvious that r does not belong to the plane of discrete points. To point and point, the relationship between the point distribution in plane is relatively dense, while points outside the plane went into a state of scattered, so set the threshold value r_0, within the scope of the radius r of the neighborhood point number if comply with the threshold value r_0 filter condition, but a continuous distribution point is judged to be plane, instead is judged to be outside the plane. This completes the on-board LIDAR point cloud data, building facade of the coarse segmentation.

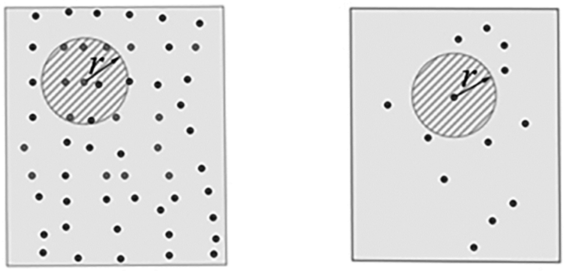

Fig. 5. Based on the point of the radius r density distribution

It may be similar to the same plane when the coarse segmentation point cloud are completed. Using the following two conditions to rest another fitting into the same plane:

(1) The difference between the origin of coordinates to the distance and the rest two another is less than set threshold.

(2) The rest two another Angle is less than the set threshold. As shown in Figure 5, D_1, D_2 from building facade extracted two point clouds, p_1 and p_2, respectively, in the rest two another points, the distance between p_1 and p_2 vector for r_{12}. Normal vector of p_1, p_2, are respectively n_1 and n_2. The two sides computing formula of angle for:

$$\kappa = \cos^{-1} \boldsymbol{n}_1 \boldsymbol{n}_2 \,. \tag{13}$$

The two sides of distance different formula can be calculated as below:

$$D = \max\left(\left|\boldsymbol{r}_{11} \cdot \boldsymbol{n}_1\right|, \left|\boldsymbol{r}_{11} \cdot \boldsymbol{n}_2\right|\right) \,. \tag{14}$$

Set distance difference threshold D_0 and κ_0 angle threshold, rest when screening another conforms to two threshold conditions at the same time, the two sides can be fitted into the same plane.

6. The experiment and analysis

On-board LIDAR system for the experimental data of the measured point cloud data respectively based on gradient difference method and the method based on projection point density compared to the experimental data analysis, purpose is to obtain the overall performance of the filtering method based on gradient difference.

Click in a certain city between blocks, in the process of data collecting, measuring car even splits at a speed of 30 km/h, at the same time GPS and INS record measurement car dynamic position and attitude parameters of each sensor to get original data acquisition of information. The original point cloud can be shown in Fig. 6.

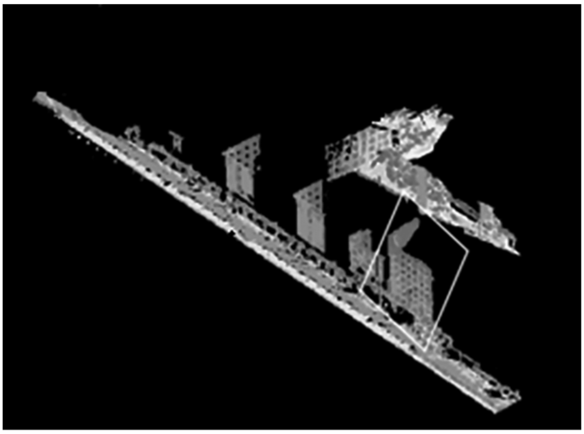

Fig. 6. Experimental blocks the original point cloud data

In order to verify the practicability of the algorithm, the typical characteristics of point cloud data is chosen from the block in experiment data, which is in Fig.6 can be shown with white line painted area. Respectively with the method based on projection point density and the method of data processing, and the results are compared.

The enlarged original point cloud picture is shown in Fig. 7. Slope difference theta of the method in the experiment is 0 to 100, threshold of parameter n_0 is 4. Processing the results as shown in Fig. 8, which black points on behalf of the building facade, another black dot represents the ground points. We can see that building facade point is basically complete. Based on the projection point density method processing and the time are needed for processing of 46 s and 20.5 s, respectively.

In order to facilitate analysis, selection of building facade, bottom line with tree, closely adjacent area such as pole processing result comparison, Fig. 9, the left side of the Figure. 10 regional respectively for two methods of processing results and its corner area enlarge figure, it is found that in Fig. 9 buildings around the tree pole, pole, such as object is difficult to eliminate, this is mainly because the tree pole, pole and zero facade with similar characteristics, it is not easy to be to get rid of. Line segment segmentation in this paper the method to eliminate the discrete point cloud, at the same time, through the distance threshold line segment classification good eliminates the lines of trees, such as pole linear target, the effect of the front of the building facade obstructions are removed effectively, is advantageous to the late point cloud of feature extraction and modeling (Fig. 10).

In order to quantitatively verify the effect of the method proposed in this paper, it is necessary to choose manual extraction method from the original data extracting

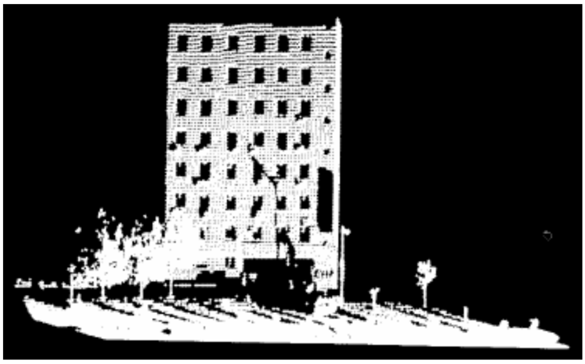

Fig. 7. Point cloud data

Fig. 8. After classifying data point cloud

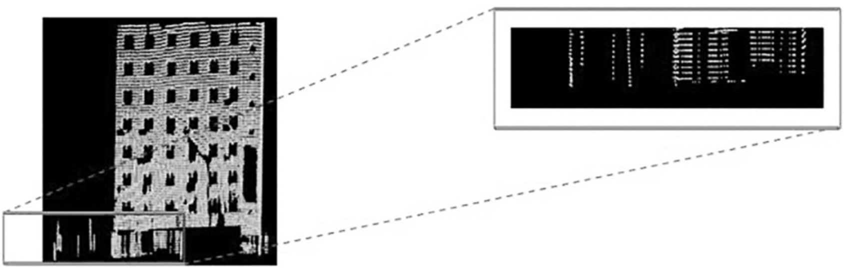

Fig. 9. The projection point method of classification result

building facade, as reference data for classification. At the same time, draw lessons from the error of the airborne LIDAR data filtering evaluation indexes, the type I error is defined as a building facade point error is divided into the building facade point error, the second type of error is defined as the building facade point error is divided into building facade point error, error analysis results are given in Table 1.

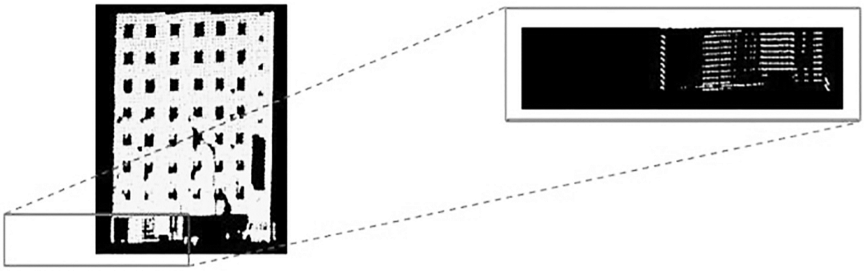

Fig. 10. The classification of the gradient difference results

Table 1. The experimental results contrast table

Manual method (points)		Projection point density method				The methods			
Facade	Other	Facade	Other	Error I	Error II	Facade	Other	Error I	Error II
16394	50327	16389	50409	6.21 %	2.09 %	16269	50395	3.59 %	0.61 %

Through the analysis of the above experimental results, the conclusion can be drawn as follows:

(1) The proposed data filtering method based on gradient difference, compared with the projection point density method, avoid a lot of the distance between the point of calculation, algorithm efficiency is improved.

(2) Data filtering processing error I and II are controlled within 10 %, show that this filtering method has higher precision.

7. Conclusion

(1) Detailed on-board LIDAR system is analyzed to obtain the spatial distribution characteristics of point cloud data, including: building facade point distribution, distribution characteristics of ground point and other distribution characteristics of feature points.

(2) Put forward based on gradient difference of on-board LIDAR point cloud data filtering method, and analyses the four steps of its implementation, including: point cloud in addition to dry, based on the slope difference of the point cloud segmentation, the classification of the small line point collection processing and data correction.

(3) The paper expounds the current several mature point cloud facade segmentation method, and mainly discusses the vehicle based on random sampling algorithm of LIDAR point cloud building facade segmentation is implemented.

(4) By introducing a real neighborhood point cloud data comparative analysis of the experiment, further demonstrates the effectiveness of the method given in this

paper.

References

[1] K. Zhang, S. C. Chen, D. Whitman, M. L. Shyu, J. H. Yan, C. C. Zhang: *A progressive morphological filter for removing nonground measurement from LIDAR data.* IEEE Transactions on Geoscience and Remote Sensing *41* (2003), No. 4, 872–882.

[2] K. Zhang, D. Whitman: *Comparison of three algorithms for filtering airborne Lidar data.* Photogrammetric Engineering & Remote Sensing *71* (2005), No. 3, 313–324.

[3] Q. Chen, P. Gong, D. Baldocchi, G. Xie: *Filtering airborne laser scanning data with morphological methods.* Photogrammetric Engineering & Remote Sensing *73* (2007), No. 2, 175–185.

[4] S. Filin, N. Pfeifer: *Neighborhood systems for airborne laser data.* Photogrammetric Engineering & Remote Sensing *71* (2005), No. 6, 743–755.

[5] X. H. Zhang: *Airborne laser scanning altimetry data filtering and feature extraction.* Wuhan University, Ph.D. Thesis (2002).

Received April 30, 2017