

Feature Extraction of Buildings Based on Vehicle-borne LiDAR Data

Lin Sheng¹, Shuangting Wang¹, Gaowei Zhou²

¹School of Surveying & Land Information Engineering, Henan Polytechnic University, Jiaozuo, China

²Chinese Academy of Surveying & Mapping, Beijing, China

Email: shenglin417@163.com

Abstract—Vehicle-borne LiDAR is a new technical means to quickly obtain high precision three-dimensional data for the city modeling and digital city construction. But the data are discrete and no topological relations. So this research is focused on the vehicular laser-scanning system to obtain the feature lines of buildings. First of all, the point cloud data of surface features are classified, and then the building feature points are extracted successfully by the way of horizontal projection and section search. Using linear regression model verifies that the feature points of the different sections can fit to characteristic line of linear significant. Under the premise of ensuring the accuracy, this method can be achieved of the extraction of the feature points of buildings and of the classification of ground objects.

Key words—vehicle-borne LiDAR, point cloud classification, feature extraction, building modeling.

I. INTRODUCTION

Laser scanning technology is a technical means to quickly obtain high precision space information. Three-dimensional laser scanning system, according to the platform., can be divided into space-borne laser scanning system, air-borne laser scanning system, vehicle-borne laser scanning system and ground laser scanning system. Vehicle-borne laser scanning systems, also known as vehicle-borne LiDAR (Light Detecting and Ranging), can get high precision three-dimensional spatial data and image information with the principle of rapid, accurate and flexible, so it has been applied in various fields, such as city modeling and road monitoring.

The feature extraction of building is the premise and key to city modeling and digital city construction. But the characteristics of vehicle-borne LiDAR data increase the difficulty of the point cloud classification and feature extraction. The main characteristic as follows: 1)When acquiring city elevation point cloud, we can collect not only the information of buildings but also some other redundant surface features, because the ground objects such as trees and utility poles always lie between the scanner and building, causing incomplete the point cloud of building. 2) Small sampling interval and a large amount of data [6]. 3) The data are discrete and no topological relations each other. As a whole, there is no fixed structure data. 4) The objects have a wide variety of features, different structure and different distribution, which is difficult to classify all features according to a fixed pattern. Due to these reasons, classification and extraction process for vehicle-borne LiDAR data is more

complex at home and abroad. There is no doubt that no methods that are simple, practical and strong adaptability, which is a bottleneck of the development of vehicle-borne laser scanning technology. So in this paper, the vehicle-borne LiDAR data classification and building extraction methods are researched to explore an effective way to solve the problem of point cloud data processing and to provide technical support for modeling of buildings and towns.

II. TECHNIQUE FLOW

It is a problem to separate building point clouds correctly from other features, so solving this problem properly must have a clear thought and strict technical process. Firstly, the filters are applied to the original point cloud data, filtering out non terrain echo noise point, disconnected points deviating from the main cloud far than connection point cloud and isolated points of background objects; Secondly, according to the surface features of the experimentation area, low - surface coarse classification model is built, and then, basing on this model, point cloud are divided into four categories: vegetation, roads, buildings and other objects (guardrail, bench and road signs etc.); Thirdly, the process for building point cloud have been carried on the extraction of the feature lines; Finally, the precision of classification and accuracy of feature points are extracted are verified by using linear regression model. In this paper, the technical process is shown in figure 1.

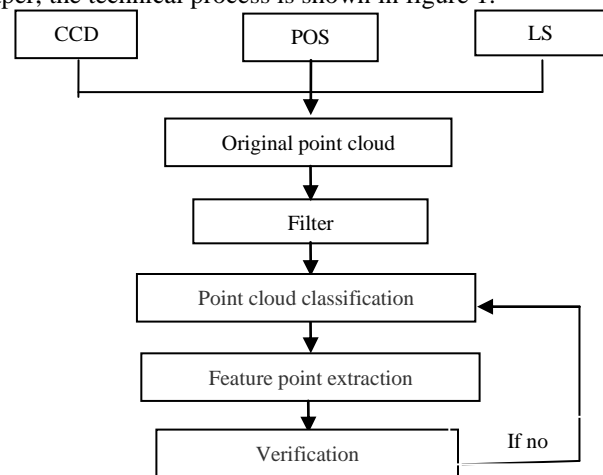


Fig.1 Technique flow

III. POINT CLOUD CLASSIFICATION AND BUILDINGS FEATURE EXTRACTION

A. Classification of point cloud

In this paper, the feature point cloud classification principle is first established low - surface model and other surface features of the city are classified by comparison with this model height difference. Low - surface model is built by Axelsson improved triangulated irregular network encryption methods, the basic principle as follows: firstly, obtain an initial sparse triangulated irregular network by minimum adjacent areas algorithm, and then add the points meeting the threshold condition to the triangle network, and rebuild the triangulated irregular network. Finally, calculate the new threshold condition. The remaining points are determined in the same manner Repeat iterations of this process until there are no new point is added so far ^[21].

The judgment method of adding a new point is: the target point to the corresponding vertex angle of the triangle and the distance of the target point to the triangular surface, respectively, compared with the set threshold value. If less than the threshold, the point is sentenced to ground point, otherwise the object point. The choice of the parameter values of the threshold condition is critical, directly determining the correctness of the added ground point. The selection process: first calculate the size of the respective angles of a triangle of irregular triangular mesh and the triangle side length, and second achieve statistical angles size and edge length value, and then determine the angle threshold value and the edge length threshold size by the value in the histogram. Note that the size of the angle threshold and edge length threshold are recalculated every time you add a new point. The irregular triangular mesh approximation of the ground is made after the calculation process ^[2]. Then, determine the point category, according to the point falls within a height range interval compared with the surface model. And then the point is divided into low vegetation, medium vegetation and high vegetation. For the building data, classification method is to create a moving window, according to the minimum housing area parameters, coinciding to the angle of inclination of the adjacent points and height difference with the input parameters, at the same time combining the shape characteristics of the building data for classification.

B. Feature extraction of buildings

The feature extraction of buildings, which provides data support for the construction of digital city and city modeling, has a strong practical value, but the feature extraction of vegetation is not much significance. On the other hand, the data structure of pavement is complex and has varied forms so that the extraction method is more difficult. Therefore the feature extraction of pavement is not discussed in this paper.

In this paper, the feature extraction of buildings is in accordance with the feature of the buildings of the vehicle-borne LiDAR data, because the façade information of the buildings is rendered as “┌┐”. So, at first, make the building information projected onto the horizontal plane turning 3D data into two-dimensional data; then extract the coordinate

values of each node and select a value as a reference of x or y(the following x as an example). Make X values of all nodes in descending order and divide into different sections range; then the grid is divided by a certain boundary value in each zone to the unit, finding out the maximum and minimum elevation point in each unit, which can be composed of the feature line of the building. Specific process is shown in Figure 2.

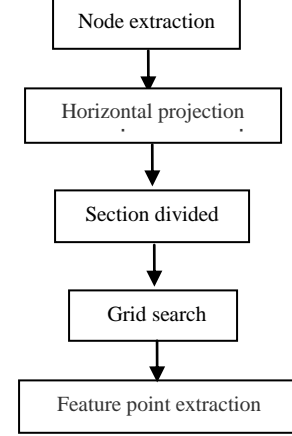


Fig.2 Flow chart of feature point extraction

C. Linear regression model

It is necessary to further verify the accuracy of the classification results and the accuracy of feature extraction of buildings. Currently, the extracted feature points are still three-dimensional data, however, each of the feature points is located in their two-dimensional feature line for each zone. So the goal of the classical linear regression model is to verify whether the feature discrete points are linearly related in each section range to achieve the purpose of verifying the accuracy of extraction and classification.

1) Model theory

Linear regression model based on the sample as follows:

$$\begin{cases} y_i = \beta_0 + \beta_1 x_i + \varepsilon_i & i = 1, 2, \dots, n \\ \varepsilon_i \in N(0, \sigma^2), i = 1, 2, \dots, n, \text{independent of each other} \end{cases} \quad (1)$$

In the formula (1), y_i is the response variable; x_i is the explanatory variable; β_0 and β_1 are the unknown parameters; β_1 is the regression coefficient; ε_i is the random error. And the linear part is $\beta_0 + \beta_1 x_i$, the random error ε_i should meet formula (1). Then using the principle of least squares makes fitting error: $(y_i - (\beta_0 + \beta_1 x_i))$ minimum, the following:

$$Q(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (2)$$

In the formula (2), $Q(\beta_0, \beta_1)$ is error sum of squares of straight-line fitting at all x_i . Obviously, the smaller the $Q(\beta_0, \beta_1)$, the better the fitting of the data and the straight line. The values of $\hat{\beta}_0$ and $\hat{\beta}_1$ making $Q(\beta_0, \beta_1)$ minimum

respectively are as both the least squares estimation [8], the following:

$$Q(\hat{\beta}_0, \hat{\beta}_1) = \min_{\beta_0, \beta_1} Q(\beta_0, \beta_1) \quad (3)$$

As following the regression equation:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (4)$$

2) Significant test of regression

The change of the response variable is generated by a linear function of the explanatory variables and random errors, so the test statistics based on the idea of hypothesis testing can be carried out. The hypotheses:

$$H_0: \beta_1 = 0, H_1: \beta_1 \neq 0 \quad (5)$$

According to significant F-test of the regression equation, the significant level α , the rejection region is:

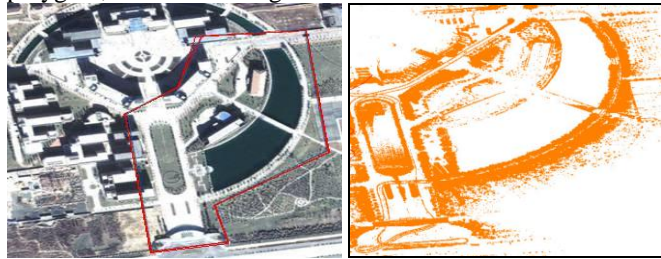
$$\left\{ F = \frac{\frac{S_R}{S_e}}{n-2} \geq F_\alpha(1, n-2) \right\} \quad (6)$$

In the formula (6), $S_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, $S_e = \sum_{i=1}^n (y_i - \hat{y}_i)^2$. If F value falls in the rejection region, then there will have a significant linear [8].

IV. THE RESULT AND ANALYSIS OF EXPERIMENT

A. The introduction of experimental area

The experimental area is a part of the campus of Henan Polytechnic University, including buildings, guardrail, stone bench, road signs, streetlights, trees, lawn and other objects. The terrain of this Experimental area is relatively flat; most of the buildings show distinct edges and corners of the polygon., as shown in Figure 3.

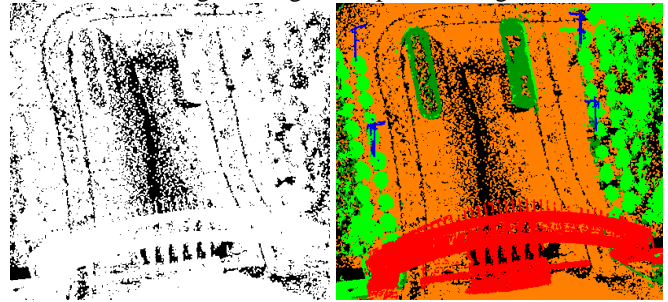


a Remote sensing image b Point cloud map
Fig.3 Experimental area

B. Classification of ground objects

According to the distribution of ground objects in experimental area, the ground objects is divided into four big classes: vegetation, roads, buildings and other ground objects. After pretreatment and filtering, the noise points and non-connection point cloud can be eliminated. According to the classification principle, the classification process is divided into coarse-classification and fine-classification. The coarse-

classification based on Low-Surface Model divide the point cloud into low point, surface and other objects. The fine-classification based on the results of the coarse-classification classifies in accordance with the Vegetation -- Building -- other objects (guardrail, bench and road signs etc.). For some objects, there are wrong leakage points, especially for buildings. It is essential to make a little artificial adjustment based on the shape characteristics of the point cloud data. The final classification is obtained as shown in Figure 4: Red represents building, orange represents pavement, blue represents streetlight, and green represents vegetation.

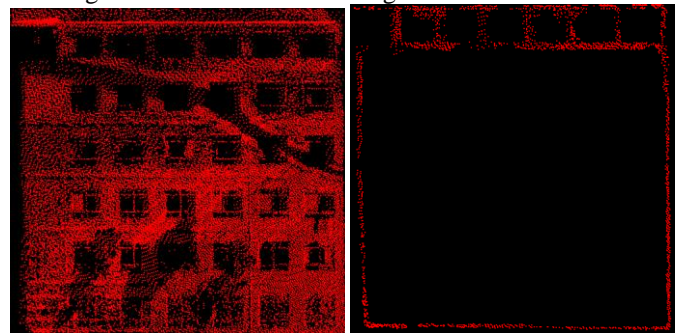


a Original point cloud b Classification

Fig.4 Classification results

C. Feature point extraction of building

At first, we extract the Building data from the classified vehicle-borne LiDAR data; second, read the data of point cloud with TerraScan, to find the node value through “show location” and export data; third, make the data format into excel format using Surfer; fourth, divide the data into different sectors using the node values in the VC++ environment, which are equivalent to different two-dimensional array. Then find out the maximum and minimum points of the elevation in each unit by making a search of the data, based on the unit of the grid of 0.1m*0.1m in the each section. And these points can be composed of the characteristic line of the building. Extracted building feature line is shown in Figure 5.



a Point cloud of building b Feature lines

Fig.5 Feature lines extraction

D. Model validation

Before Linear Regression Model validation, the data are simplified using Surfer to make three-dimensional data into two-dimensional data. Two coordinates of two-dimensional data are expressed by V1 and V2: V1 is the response variable and V2 is the explanatory variable. There are a number of

assumptions in the establishment of model. So the processed data must meet the requirements of homogeneity of variance and residual normality, therefore it needs to determine whether the building data meet the above two requirements through standardized residual、predicted values as well as the relationship between the standardized residual frequency. The test results as shown in Figure 6 and Figure 7.

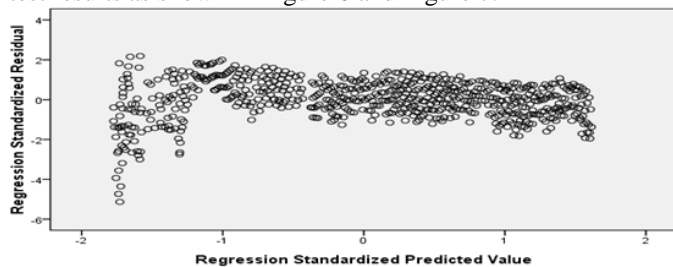


Fig.6 Scatter plot

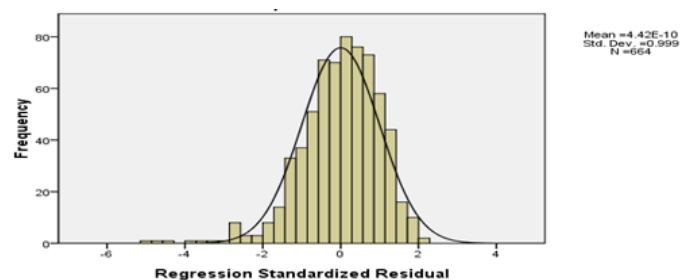


Fig.7 Histogram of standardized residuals

Figure 6 shows the scatter diagram: regression predictive value is X-axis and the standardized residual is Y-axis. Obviously, the standardized residuals are basically between -2 and 2, and are consistent with the requirements of homogeneity of variance. Figure 7 shows histogram of standardized residuals: regression standardized residuals is X-axis, and the standardized residuals frequency is Y-axis. The residual distribution of a normal distribution indicates that Building data satisfy linear regression model and can carry out regression analysis. The buildings feature point data are analyzed by Linear regression using the Formula (1) ~ Formula (4). The results are shown in Table 1.

Tab.1 Model Summary

Mod	R	R Square	Adjusted R Square	Std. Error of the el
1	.903 ^a	.815	.712	.10213744348893

a. Predictors: (Constant), V2
b. Dependent Variable: V1

In the table, R represents a multiple correlation coefficient which indicates the degree of the linear relationship of the explanatory variables on the response variable. The larger the value is the more significant the linear relationship. It can be seen from Tab.1: the value is larger than 80%, showing good linearity.

In accordance with the regression model and the principle of the test of significance, ANOVA is shown in the Tab.2.

In this table, df represents the degrees of freedom, F represents the F-test value of significant analysis: the higher the value of F, the greater the impact of independent variables on the dependent variable, the smaller the outside influence. Sig represents the probability value corresponding to F value. Sig which is less than 0.01 also shows significant linear. Linear regression analysis verify the characteristic line of extracted buildings feature points as a straight line and edge features of actual buildings are consistent.

Tab.2 ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.857	1	4.857	465.6	.000 ^a
	Residual	6.906	662	.010		
	Total	11.763	663			

a. Predictors: (Constant), V2 b. Dependent Variable: V1

V. CoNCLUSION

In this paper, the building feature points are extracted successfully by the way of horizontal projection and section search for point cloud of experimental area. Finally, using linear regression model verifies that the feature points of the different sections can fit to characteristic line of linear significant. Under the premise of ensuring the accuracy, this method can be achieved of the extraction of the feature points of buildings and of the classification of ground objects, to better provide some technical support for the field of urban modeling and road monitoring.

REFERENCES

- [1] F.F. Wu and Q.Q. Li, "Based on vehicle-borne laser scanning data of target classification method," Science of Surveying and Mapping, vol. 32, pp. 75-77, Jul. 2007.
- [2] J.B. Tao, "MicroStation user guide," CA: Beijing, 1993, pp.1-58.
- [3] X.S. Lu, and Q.Q. Li, "Vehicular city information acquisition and 3d modeling system," Journal of Wuhan University, vol. 4, pp.76-80, 2003.
- [4] W.ZH. Shi and B.J. Li, "Based on the projection point density of vehicle-borne laser scanning distance image segmentation method," Science of Surveying and Mapping, vol.34, pp. 95-100, 2005.
- [5] Y. Yang, Y.SH. Zhang, "Based on scan line of vehicle-borne LiDAR point cloud filter method," Journal of Surveying and Mapping Science and Technology, vol. 27, pp. 209-212, 2010.
- [6] F. X. Guan, F.W. Li, "LiDAR data characteristics and classification algorithm," Journal of Surveying and Mapping and Spatial Geographic Information, vol. 32, pp. 81-82, 2009.
- [7] ZH.ZH. Zhang, X.ZH. Xu, Application of mathematical statistics, CA: Beijing, 2008, pp.168-179.
- [8] K.Z. Yu, "Linear regression analysis using SPSS," Journal of Chongqing Institute of Technology, vol.16, pp.29-34, Apr, 2002.