

# Research on Building Facade Point Cloud Data Based on Random Sampling Algorithm Segmentation

Peng Cang<sup>1,a</sup>, Zhenglin Yu<sup>1</sup>

<sup>1</sup> Changchun University of Science and Technology ,Changchun,130022,China;

<sup>a</sup>[yubo745@163.com](mailto:yubo745@163.com)

**Keywords:** Building point cloud; Division of the facade; Random sampling

**Abstract:** The point cloud data for building facade segmentation problem, based on random sampling algorithm is proposed building facade segmentation method, point cloud data, and analyzes the principle and operation process of the algorithm, at the same time, this paper expounds the key parameters in the process of algorithm implementation basis set. Through experimental analysis in the paper, further verify the effectiveness of the method given in this paper.

## Introduction

Nowadays, building facade point cloud segmentation method mainly has: the point cloud segmentation method based on boundary<sup>[1]</sup>, point cloud segmentation method based on clustering<sup>[2]</sup>, region growing method<sup>[3]</sup>and point cloud segmentation method based on random sampling consistency algorithm.

Considering the random sampling algorithm good processing ability, even if the original data contains more than fifty percent of the abnormal data, through algorithm processing will still be able to get ideal results. This paper will use random sampling algorithm, set the corresponding judgment parameters on after classification of vehicular laser radar elevation point cloud data segmentation.

## Algorithm principle

Random sampling algorithm consistency assumption, given a set of data can be described by mathematical model, can find a mathematical model of this group of data to calculate the parameters of the method. Consistency by using random sampling algorithm based on point cloud model segmentation, first of all, from the original point cloud data in constant search optimal model of point cloud model and gradually reduce the number of iterations. Then , randomly selected from the best model of a point as initial points, calculate the selection point in the model parameters. Finally setting the threshold value for all points in the initial point cloud data filtering, can obtain belongs to all the points of the model. So many times, each time to get the results compared with the last time the result, if the result is better than a result, instead of the previous model parameters, and set the number of iterations is complete. The algorithm includes six kinds of input parameters: Distance threshold  $t$ 、 Determine the minimum points of the model  $N$ 、 Points of the same model contains the largest  $n$ 、 For at least one of the best in the multiple sampling observation of minimum probability  $P_0$ 、 original data error rate  $\eta$  And the minimum number of iterations  $K$ .

### consistency random sampling algorithm specific processes

- (1)First, put the original point cloud point set  $M$ , Calculate the input parameters.
- (2)Extracted from the original point cloud to determine  $n$  points of the model, A set of points  $S$ .

(3) Judge point set M and each point in point set S I set T meets the distance threshold T. If the distance is less than the distance threshold value to model t, The will that point and create S midpoint as new sets S1.

(4) If the number of point set S1 midpoint set threshold value greater than or equal to N, Argues that the model parameters, correct record for the model Maybe\_A.

(5) Circulation step 2 to 4, retain the best model parameters.

(6) Determine whether the number of iterations is equal to the set parameters K, If is equal to K and plane Maybe\_A meet the threshold limit of N, Regarded as the best model Best\_A. If the model Maybe\_A does not comply with the threshold limit of N, As operation failed.

**parameter setting**

(1) the maximum number of iterations K

To capture at least one of the best in the K times sampling observation of minimum probability P of computation formula is as follows:

$$P_0 = 1 - (1 - (1 - \eta)^n)^K \tag{1}$$

In the type(1):  $\eta$  is the original data error rate, n is extracted from the original data can determine the minimum points of the model. The differences between the  $\eta$  is usually according to different raw data, Concrete numerical value based on the original data set; n is extracted according to the specific model, For example, a plane would take at least three points can be determined. In general, to extract a complete building facade need at least 200 points. To transform type (1) :

$$1 - P_0 = (1 - (1 - \eta)^n)^K \tag{2}$$

The type (2) the exponential:

$$K = \frac{\log(1 - P_0)}{\log(1 - (1 - \eta)^n)} \tag{3}$$

Suppose P0 was 99%, The minimum number of iterations and threshold value of n and K raw data error rate  $\eta$  relations are shown in table 1 below.

Table 1. In the n、K and  $\eta$  value under the influence of the parameters

The minimum points of the model n	The original data error rate $\eta$ (%)				
	0.2	0.3	0.4	0.5	0.6
n=3	6	11	19	34	70
n=4	9	17	33	71	178
n=5	12	25	57	145	447
n=6	15	37	96	292	1122

According to table 1, we can draw different n values are shown in figure 1 below the number of iterations increases with error rate and produce corresponding change. Known from the figure, If you want to get more ideal model, requires the higher the number of iterations.

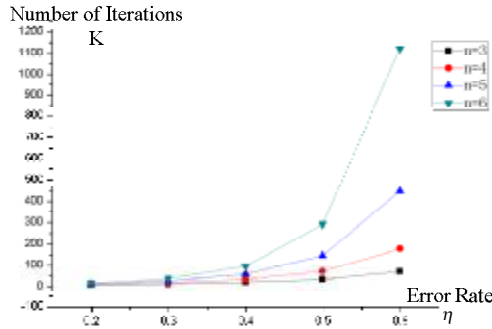


Figure 1. The curve of n, K and η

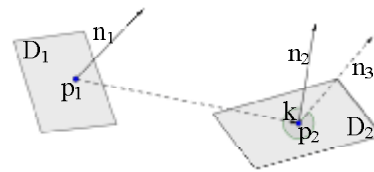


Figure 2. Angle and distance constraints

(2) The distance threshold t

Usually the distance threshold t set based on the mathematical model of actual extraction. First plane parameter mathematical model is set up, for most of the buildings, it has a simple, rules of geometry and building Angle is 90 °, so its contours in 3 d space coordinate system X axis and Y axis direction perpendicular to each other. Usually, in the same plane point should satisfy:

$$Ax + By + Cz = D \tag{4}$$

Among them, (x,y,z) is the plane three-dimensional space coordinates, (A,B,C) is the plane unit normal vector and the sum of squares is equal to 1, D is the origin of coordinates to the distance of the plane. From the classification of point cloud data to extract 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. Assuming that the points on the building façade S=(X<sub>N</sub>,Y<sub>N</sub>,Z<sub>N</sub>), N is the total number of points in the collection of all, in the form of a matrix is expressed as:

$$\begin{bmatrix} X_N & Y_N & Z_N \end{bmatrix} F = 0 \tag{5}$$

The basic matrix F is:

$$F = [a \ b \ c \ d]^T \tag{6}$$

Basic matrix, therefore, there are three degrees of freedom, choose three points as a plane point, calculate the plane parameters, according to the plane parameters to calculate the other point in the plane. Instruction in P = (X, Y, Z), plane (P1 = (a, b, c, d) P to plane Euclidean distance for P1:

$$D(P, P_1) = |ax + by + cz - D| \tag{7}$$

Normally, plane of the distance between point and plane is 0, but what we call the plane is actually a collection of points is not the real plane, so by setting distance threshold t synthesize the approximate fitting plane. In building facade point cloud data, and the plane distance if is greater than the set threshold t, due to the decision point for the out-of-plane points, judged to be plane conversely. Considering building its own structure, the complexity and the detail characteristics, distance threshold t set according to the actual situation, in order to guarantee the late detail feature extraction and modeling, the more complex structures, the distance threshold t set stricter. Figure 3 schematic for distance threshold t.

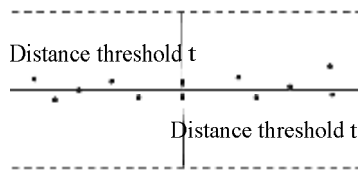


Figure 3. The distance threshold t

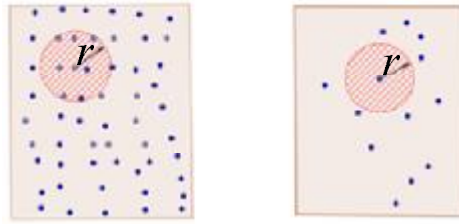


Figure 4. Based on the point of the radius r density distribution

(3) Refer to the threshold value  $r_0$

If just rely on distance threshold t to filter, then cannot fully guarantee the accuracy of segmentation. So using graphics r radius density in principle to screen out of misjudgment point again, figure 4 is based on the point of the radius r density distribution. Seen from the figure, with the radius r of the circle, the obvious do 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 conform to the threshold value  $r_0$  filter and a continuous distribution point is judged to be plane, instead is judged to be outside the plane.

On the basis of the above parameters setting, consistency using random sampling algorithm can complete the coarse segmentation of building facade.

**Patch fitting**

Because only rest get after coarse segmentation of point cloud another approximation in the same plane, not all in the same plane, so the use of the following two conditions, the fitting into the same plane. That is:

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

(2) the rest two another Angle is less than the set threshold. As shown in figure 2,  $D_1, D_2$  from building facade extracted two point clouds,  $p_1$  and  $p_2$  respectively in the rest two another point, the distance between  $p_1$  and  $p_2$  vector for  $r_{12}$ . The normal vector of  $n_1$  respectively  $p_1, p_2, n_2$ .

An Angle from the two sides figure 4 piece:

$$\kappa = \cos^{-1} \vec{n}_1 \cdot \vec{n}_2 \tag{8}$$

Origin of coordinates to the rest of another poor distance D as follows:

$$D = \max(|\vec{r}_{11} \cdot \vec{n}_1|, |\vec{r}_{11} \cdot \vec{n}_2|) \tag{9}$$

Set distance difference threshold  $D_0$  and Angle threshold  $k_0$ , when screening and conform to the rest of another two threshold conditions, the rest will be another fitting into the same plane. Rest after another fitting after get the final LIDAR point cloud building facade segmentation result.

**Experimental analysis**

Choose a set of point cloud data after filtering, the building facade (see figure 5), the data contained in the rest three features another: building owner wall, side wall and the building Windows. Consistency algorithm using random sampling on the facade segmentation experiments. Distance between point and point to 5 cm, Error rate  $\eta$  is 0.2, After calculated to obtain the best model probability  $P_0$  for 99% of the cases, the minimum number of iterations K is 1000. Set the distance threshold t as 0.03 m, Circle radius r is 1 m, refer to the threshold value  $r_0$  is 10.

Experiment result is shown in figure 6 and figure in different colors represent different, better segmentation algorithm.



Figure 5. The original building facade point cloud



Figure 6. The segmentation result

## **conclusion**

(1) Based on random sampling algorithm was analyzed and the realization of the point cloud data of building facade method principle and the concrete operation process.

(2) This paper expounds the building elevation point cloud data based on random sampling algorithm segmentation method of key parameters in the process of its implementation set criteria.

(3) Through the experimental analysis, based on random sampling algorithm is given in this paper is verified the effectiveness of the point cloud data of building facade segmentation method.

## **Reference**

- [1] Zhang K, Dean Whitman. Comparison of Three Algorithms for Filtering Airborne Lidar Data[J]. *Photogrammetric Engineering & Remote Sensing*, 2005, 71(3): 313–324.
- [2] Hu Y, Tao C V. Hierarchical Recovery of Digital Terrain Models from Single and Multiple Return Lidar Data[J]. *Photogrammetry Engineering & Remote Sensing*. 2005, 71: 425-433.
- [3] Sithole G. Filtering of laser altimetry data using a slope adaptive filter[C]. *International Archives of Photo grammetry and Remote Sensing*, 2001, 34(3/W4): 203-210 (Annapolis, ML).