Research on Automatic Denoising of LIDAR Point Cloud Data for Substation Equipment Based on Spatial Grid Density

Xinle Yu\textsuperscript{1,*}, Yong Du\textsuperscript{2}, Hao Wang\textsuperscript{2}, Jingsong Yao\textsuperscript{2}, Yuandong Wang\textsuperscript{3} and Xiaojun Shen\textsuperscript{1,}\textsuperscript{*}

\textsuperscript{1}Department of Electrical Engineering, Tongji University, Shanghai 201804, China
\textsuperscript{2}State Grid Hubei Electric Power Co., Ltd. maintenance company, Wuhan 430050, China
\textsuperscript{3}State Grid Inner Mongolia East Electric Power Co., Ltd. maintenance branch, Tongliao 028114 China

*Corresponding author. Email: xjshen79@163.com

Abstract. Aiming at the shortcomings of the existing 3D point cloud data automatic extraction methods of substation equipment, which are highly dependent on big data algorithms and low efficiency, this paper proposes a 3D LIDAR point cloud data segmentation method and process based on the multidimensional subspace grid density difference. The proposed method is based on eliminating the flying spots of 3D point cloud data, and is divided into equipment point cloud data and ground point cloud data based on point cloud data characteristics for 3D real-world modeling and accurate positioning of the model; Among them, the equipment point cloud data uses a multi-dimensional density difference segmentation method. The long-distance terrain is divided in the $XOY$ and $YOZ$ planes, and converted into a combination of multiple small-scale scale spaces. Effective segmentation, so that automatic extraction of substation equipment can be realized; The ground point cloud data uses a single-dimensional density difference segmentation method to dilute the ground point cloud data to obtain clear positioning points. The feasibility verification results of cloud data of a UHV substation show that the proposed method can effectively suppress the noise interference of interference points, realize accurate extraction and location of substation equipment, and the algorithm has high efficiency and strong engineering application.

Keywords: substation, LIDAR point cloud data, automatic extraction.

1. Introduction
As a more efficient technology for substation operation and maintenance, visualization of 3D information in substations has gradually become the focus of research. Reconstruction of 3D real-world models of substations is the technical basis for 3D visualization of substations. High-precision model reconstruction is the key to achieving a clearer and more intuitive 3D visualization. At present, the methods of 3D real-world modeling of substations mainly include VRML modeling, geometric modeling, and point cloud data-based modeling methods [1]. The first two methods have lower modeling accuracy and efficiency, while the point cloud data-based modeling method uses ground-based 3D laser scanning technology and adopts non-contact measurement and acquisition
technology to accurately collect 3D point cloud data of substation equipment and connecting lines. In practice, it can effectively avoid the shortage of traditional measurement methods, and it has gradually become the development trend of 3D real-world modeling of substations [2,3].

Substation equipment modeling technology which bases on ground LIDAR can generate massive point cloud data. The fast and efficient implementation of sub-area and point cloud data segmentation and extraction of substations can effectively improve the timeliness and engineering of the technology Application value [4,5]. Literature analysis results show that although automatic extraction of substation equipment has received some attention, related research is still in its infancy. Carlos A Vanegas [6] and others proposed an effective point automatic segmentation and extraction method for large-scale 3D point cloud data of Manhattan-type buildings, but most of its applicable objects were limited to objects with relatively single shapes and few special shapes. Wang Hanyun [7] Used the hidden shape model to describe objects, and used the Hough forest architecture to perform block detection on urban point cloud data images, and proposed circular voting and distance weighted voting mechanisms, but there were problems of complicated steps and low efficiency. Dou Benjun [8] and others proposed a method for 3D point cloud data identification of substation equipment based on improved neural network algorithm, Fang Yanjun [9] and others proposed a 3D point cloud data classification method for substations based on random forest method, and Li Lingxia [10] and others proposed The three-dimensional point cloud data classification methods based on particle swarm optimization for substations are presented. However, these methods are relatively dependent on big data processing methods and cannot maintain high efficiency and accuracy under complex external conditions. There are certain limitations in practical modeling use.

The amount of 3D point cloud data of electrical equipment is extremely huge. There are about tens of thousands to hundreds of thousands of points for one device. However, when we conducted preliminary research and analysis on the reconstruction of 3D simulation models of substations, the extraction process is identified by human eyes and processed by specific software manually, which not only requires lots of manpower and time, but also has low efficiency and processing accuracy. Therefore, it is urgent to find out how to quickly and efficiently process point cloud data, use computers instead of people to automatically extract point cloud data of electrical equipment, identify the point cloud data and noise point cloud data corresponding to the target device and perform classification removal.

In view of this, this paper proposes a multi-dimensional subspace grid density difference segmentation method based on the 3D laser point cloud data based on the structure characteristics of the ground point cloud data of the substation, which realizes the fast and efficient automation of the equipment point cloud data to reduce the modeling cost and shorten the work cycle.

2. Point cloud data characteristics of substation equipment
i) Characteristics of ground point cloud data distribution

Ground point cloud data are concentrated in the area with the smallest elevation in vertical space, and continuously in the entire planar area in horizontal space. Actual engineering uses sparse ground point cloud data to locate the equipment. However, because the density of the ground point cloud data collected is much greater than the point cloud data above the ground, very dense ground point cloud data will block the equipment point cloud data, making it impossible to visually observe the equipment point cloud data. The situation, as shown in Figure 1a. Therefore, further local zoom operations are required to complete the device positioning, which will affect the accuracy of the positioning. At the same time, the large number of point cloud data will occupy a large amount of memory when imported into the software for processing, resulting in slower processor processing speed and Causes the software to freeze when processing operations such as point cloud data perspective changes, which greatly affects work efficiency and increases work time.

ii) Characteristics of equipment point cloud data distribution

In the vertical space, the point cloud data of electrical equipment is distributed in an area where the elevation is bigger than the ground point, and the distribution range is wide. The horizontal space has the characteristic of being basically concentrated in a small rectangular area after horizontal projection.
When the device is finely scanned, the distance between the scanning instrument and the target device is relatively short, and the scanning time is relatively long. Therefore, the point cloud data density of the device is relatively large locally, but is generally much smaller than the ground point cloud data.

**iii) Characteristics of noise point cloud data distribution**

Disturbance noise point cloud data are distributed discretely in vertical and horizontal spaces. They have a large distribution span in elevation, and they are distributed in the gaps between the point cloud data of the equipment and in the external space. The noise point cloud data has the feature of relatively sparse density. When the ground point cloud data with an extremely large amount of data is removed, although the point cloud data density of part of the device is still denser than that of the noise point cloud data, from the perspective of relative error, which locate in the surroundings of the target device and the internal gap do not belong to the device, the noise point cloud data has a large interference with the point cloud data of the actual electrical equipment data in the process of fine modeling of the equipment, that is, the density of the noise point cloud data is relatively large, as shown in Figure 1b.

![Figure 1. Initial point cloud data. (a) LiDAR point cloud data of substation; (b) Diagram of equipment point cloud data distribution characteristics.](image)

### 3. Automatic extraction algorithm for substation equipment

#### 3.1. LiDAR Point cloud data preprocessing

Cloud data from substations obtained through 3D LIDAR scanning cannot be directly used for automatic extraction of equipment. Therefore, data preprocessing should be performed before the main steps of automatic equipment extraction, including steps such as flying spot removal and data reduction.

**i) Flying spot removal**

During the original data collection process, some noise points that are far different from the equipment point cloud data in elevation and horizontal space may be generated, which are called 'flying spots'. The generation of flying spots will make the subsequent point cloud data type division and the automatic extraction accuracy of substation equipment worse, resulting in lower efficiency. Therefore, after obtaining the original data, we need to manually remove those noise points that are very obvious in the elevation and space, that is, flying point removal.

**ii) Point cloud data streamlining**

Since the original point cloud data obtained is massive, it is necessary to reduce the data while maintaining the surface structure and the accuracy that meets the conditions. The method is average reduction, that is, one point is reserved for every N points in the origin, and so on, and the original data can be traversed to complete the preprocessing.
3.2. LiDAR point cloud data segmentation method based on multi-dimensional subspace grid density difference

3.2.1. Point cloud data division. According to the analysis in the previous chapter and the analysis of the actual environment of the substation, it can be seen that the ground point cloud data has a significant difference in distribution of elevation in vertical space: In a certain range, the ground point cloud data has a certain range in elevation, that is, the elevation is relatively concentrated at the bottom of the vertical space, and there are no disturbing feature points in the space. Therefore, a simplified elevation threshold segmentation algorithm can be used to divide the point cloud data type. The algorithm works as follows:

First, input all point cloud data. The first step is to find the boundary value of the entire original point cloud data and determine the initial point cloud data distribution space $M_0$.

$$
\begin{align*}
    x_{\min} &= \min(x) \\
    x_{\max} &= \max(x) \\
    y_{\min} &= \min(y) \\
    y_{\max} &= \max(y) \\
    z_{\min} &= \min(z) \\
    z_{\max} &= \max(z)
\end{align*}
$$  \hspace{1cm} (1)

where $(x, y, z)$ is the spatial coordinate value of any point cloud data, select a suitable division scale $d_x$, and divide the entire point cloud data space into $N$ subspaces along the X coordinate,

$$
n = \frac{x_{\max} - x_{\min}}{d_x} + 1
$$  \hspace{1cm} (2)

In the formula, $|x_{\max} - x_{\min}|$ means to find the largest integer not greater than $(x_{\max} - x_{\min})$.

Then apply formula (3) to calculate the elevation difference $\delta_i$ between the point in each subspace $M_i$ ($i = 1, 2, ..., N$) and the lowest point, and use $\delta_i$ as the feature value.

$$
\delta_i = z_i - z_{\min}
$$  \hspace{1cm} (3)

Finally, set the feature value $\delta_0 = 30$ (cm). If $\delta_i \leq \delta_0$, the point $Z$ is the ground point cloud data containing device location information, and is the class A point cloud data, otherwise it is the class B point cloud data with noisy points.

3.2.2. Automatic point cloud data extraction method for substation equipment. i) One-dimensional density difference algorithm for XOY plane subspace partition

Based on the point cloud data division, the obtained equipment point cloud data still contains the interfering noise point cloud data and measurement error points. In order to extract the accurate point cloud data of electrical equipment, the noise point cloud data removal operation needs to be performed.

The point cloud data characteristic analysis results show that the point cloud data density of the equipment part is denser than that of the noisy point cloud data. Therefore, this paper proposes a multi-dimensional subspace grid density difference segmentation method. The algorithm works as follows:

First, project the three-dimensional space point of the device to the XOY plane to obtain the initial space $M_0$, and then locate the boundary of it, select the appropriate division scale $d_x$, $d_y$, and divide the initial space $M_0$ into $m \times n$ subspaces $M_{i,j}$ ($i = 1, 2, ..., m, j = 1, 2, ..., n$) along the X and Y coordinate, respectively, the formula is as follows:
\[
m = \left\lfloor \frac{x_{\text{max}} - x_{\text{min}}}{d_x} \right\rfloor + 1
\]
\[
n = \left\lfloor \frac{y_{\text{max}} - y_{\text{min}}}{d_y} \right\rfloor + 1
\]

Then, the number of point cloud data \( C_{i,j} \) in each subspace \( M_{i,j} \) is counted, and \( C_{i,j} \) is used as the feature value of the subspace.

The threshold \( C_0 \) is reasonably set according to the subspace and point cloud data density, and all subspaces are divided into two categories based on \( C_0 \). Points smaller than the threshold are judged as noise points and removed, and those larger than the threshold are equipment and retained.

It should be noted that due to the differences in the scanner model and the size of the target point cloud data itself, the threshold \( C_0 \) should be set reasonably based on the actual point cloud data quality and subspace size: When the subspace size is large and the point cloud data density is high, a larger threshold value needs to be set; otherwise, the threshold value can be appropriately reduced.

However, the one-dimensional density difference algorithm is to project the initial point cloud data onto the \( XOY \) plane. In this process, the device point cloud data \( Z \) coordinate information is lost, that is, when the \( X \) and \( Y \) coordinate attributes of the noise point in the device point and the gap are the same, When the \( Z \)-coordinates are different, the one-dimensional algorithm cannot accurately remove the gap noise, which causes a large extraction error. Therefore, this paper proposes a multi-dimensional density difference algorithm based on the one-dimensional density difference.

ii) Two-dimensional density difference algorithm for dividing \( YOZ \) plane subspace

Use the result point cloud data obtained in the previous step as the initial space \( M_0 \) of the two-dimensional density difference algorithm, and then select the appropriate division scales \( d_x, d_y \), and divide the initial space \( M_0 \) into \( m \times n \) subspaces \( M_{i,j} \) (\( i = 1, 2, \ldots, m \), \( j = 1, 2, \ldots, n \)), the formula is as follows:

\[
\begin{align*}
m &= \left\lfloor \frac{y_{\text{max}} - y_{\text{min}}}{d_y} \right\rfloor + 1 \\
n &= \left\lfloor \frac{z_{\text{max}} - z_{\text{min}}}{d_z} \right\rfloor + 1
\end{align*}
\]

Then count the number of point cloud data \( C_{i,j} \) in each subspace \( M_{i,j} \) again, and use \( C_{i,j} \) as the feature value of the subspace. The threshold \( C_0 \) is reasonably set according to the subspace and point cloud data density, and all subspaces are divided into two categories according to \( C_0 \). Points smaller than the threshold are judged as noise points and removed, and those larger than the threshold are equipment points, which are retained, that is, it can complete the segmentation of equipment point cloud data and interference noise point cloud data, and extract the electrical equipment point cloud data.

3.2.3. Ground LIDAR point cloud data processing algorithm based on single-dimensional subspace density difference.

After performing ground point recognition in the foregoing, the Class A point cloud data obtained is the ground point cloud data. The analysis result of point cloud data characteristics shows that the very dense ground point cloud data will block the equipment point cloud data, making it impossible to observe the point cloud data of the equipment. At the same time, it causes software freeze, which affects the accuracy and efficiency of positioning. Therefore, it is necessary to thin the ground point cloud data. This paper uses an improved single-dimensional subspace density difference algorithm. The principle is as follows:

First, the ground point cloud data is projected onto the \( XOY \) plane to obtain the initial space \( M_0 \), and then the boundary positioning is performed. Select an appropriate division scale \( d_x, d_y \), and divide the
initial space $M_0$ into $m \times n$ subspaces $M_{i,j}$ ($i = 1.2 \ldots m, j = 1.2 \ldots n$) along the $X$ and $Y$ coordinate. The formula is as follows:

$$m = \left\lceil \frac{x_{\text{max}} - x_{\text{min}}}{d_x} \right\rceil + 1$$

$$n = \left\lceil \frac{y_{\text{max}} - y_{\text{min}}}{d_y} \right\rceil + 1$$

(6)

Count the number of point cloud data $C_{i,j}$ in each subspace $M_{i,j}$, and use $C_{i,j}$ as the feature value of the subspace. Different from the previous chapters, the scale $d$ is relatively large when processing the ground point cloud data, and then a threshold $C_i$ is set reasonably according to the subspace and point cloud data density, and the number of point cloud data is reasonably simplified based on this threshold $C_i$ and equation (7):

$$C_{i,j} \leq C_i$$

(7)

That is, the number of subspace point cloud data smaller than the threshold is retained, which is the sparse point cloud data obtained after thinning.

4. Feasibility test and analysis discussion

4.1. Algorithm automatic extraction effect test

In order to verify the effectiveness of the automatic point cloud data extraction algorithm for the UHV substation equipment proposed in this paper, some data of the point cloud data in the DC area of a 500kV converter station measured by ground LIDAR was selected as the test object, as shown in Figure 2.

**Figure 2.** Selected excerpt data. (a) DC area point cloud data; (b) DC-filter point cloud data.

First, we perform preprocessing and point cloud data type division to obtain equipment point cloud data and ground point cloud data, respectively. We use equipment multi-dimensional subspace mesh density difference segmentation method to process equipment point cloud data, and obtain point cloud data after one-dimensional segmentation. The effect is shown in Figure 4a. It can be seen from Figure 3a that this method can effectively eliminate noise in the point-free area of the device without cloud in the $XOY$ plane, leaving only a small part of the noise in the device gap not completely removed. The device point cloud data obtained after 2D segmentation is shown in Figure 3b. It can be obtained from Figure 3b that the noise point cloud data in the equipment gap of the substation are effectively removed, the equipment point cloud data are accurately and efficiently extracted, and the combined effect of the ground point cloud data and the equipment point cloud data after the ground single-dimensional
segmentation is shown in Figure 3c, it can be seen that the ground point cloud data is effectively sparse, which greatly improves the observation effect.

![Figure 3]  
(a)  
(b)  
(c)

**Figure 3.** Diagram of equipment, ground point cloud data extraction. (a) Point cloud data extraction results from one-dimensional equipment; (b) Point cloud data extraction results from two-dimensional equipment; (c) Ground point thinning results.

4.2. *Comparative analysis of algorithm extraction accuracy and extraction efficiency*

In order to accurately test the efficiency of the algorithm and verify the reliability of the algorithm proposed in this article, we performed quantitative verification on each step of the experiment, and then manually reviewed the extracted point cloud data of the substation equipment, that is, we manually deleted the non-equipment points. Then we count the number of valid points left in the substation, and use the ratio of the number of valid points to the number of extracted points as the algorithm to extract the correct rate of the cloud at the substation. The statistical results are shown in Table 1.

In addition, the size of the ground point cloud data is 23.222 MB, the number of data points is 487714, and the number of point cloud data points obtained after one-dimensional segmentation is 23707, the size is 1.12 MB, and the thinning rate is 4.86%.

As can be seen from Table 1, the method proposed in this paper extracts more point cloud data of substation equipment than artificially extracted point cloud data. This is a small amount of noise omission caused by the equipment's point cloud data projection that does not fully form an orthographic projection, but the algorithm accuracy rate is above 98%. The algorithm results fully meet the engineering application conditions.
Of course, the theoretical noise may interfere with the accuracy of the subsequent 3D real-world modeling, and the accuracy of the extraction can be further improved by the method of projection angle correction in the future. The observation effect of the thinned floor after stitching has been improved. At the same time, from the perspective of the accuracy of the positioning operation, the algorithm in this paper retains the point cloud data of the device part relatively completely, that is, the position information of the device's point cloud data positioning is accurately retained. Based on this, and then using the ground point cloud data information as auxiliary information, we can complete the entire process of automatic extraction of point cloud data for substation equipment.

| Point cloud data type | Point cloud data size /MB | extracted point clouds /number | Correct rate /% |
|-----------------------|---------------------------|-------------------------------|----------------|
| Initial point cloud data | 81.989 | 1421592 | |
| One-dimensional density difference point cloud data | 34.287 | 702194 | |
| Two-dimensional density difference point cloud data | 21.232 | 699889 | |
| Manual review point cloud data | 21.190 | 688505 | 98.37 |

In summary, the point cloud data extraction method for substation equipment proposed in this paper can efficiently and accurately extract the point cloud data of electrical equipment in practical applications, and can effectively solve the problem that the existing methods rely too much on big data methods, and has strong engineering practicability.

5. Conclusion
Aiming at the shortcomings of the existing methods that rely heavily on big data technology, this paper proposes a multi-dimensional subspace grid density difference segmentation method based on 3D LIDAR point cloud data. This method is suitable for UHV substations and can efficiently and accurately remove and process noisy 3D point cloud data according to different scanning methods and data types, such as the difference in density under fine scanning and positioning conditions and the effect of noise. The method proposed in this paper can ensure the reliability and applicability of 3D model reconstruction of massive point cloud data. This paper performs denoising processing on massive 3D point cloud data. It can accurately and efficiently realize the automatic extraction of point cloud data of substation equipment, which can lay the foundation for reducing the point cloud data-based substation real-world modeling costs and shortening the modeling cycle.

References
[1] Wang Rende, Du Yong, Shen Xiaojun. Status and prospect of 3D modeling methods for substations [J]. North China Electric Power Technology, 2015 (02): 19-23.
[2] Zhang Rui, Du Yong, Wang Qi, et al. Research on 3D Reconstruction Method of Substation Based on Real Site Cloud Data [J]. Hubei Electric Power, 2018, 42 (01): 28-34.
[3] P. Forman and I. Parry, "Rapid data collection at major incident scenes using three dimensional laser scanning techniques," Security Technology, 2001 IEEE 35th International Carnahan Conference on, London, 2001, pp. 60-67.
[4] Shen Xiaojun, Qin Chuan, Du Yong. An automatic extraction algorithm of high voltage transmission lines from airborne LIDAR point cloud data [J], Turkish Journal Of Electrical Engineering And Computer Sciences, 2018, 26(4): 2043-2055.
[5] Fang Yanjun, Tang Mian, Luo Jia, et al. Automatic point cloud segmentation of 3D modeling of substation based on random forest method [J]. Manufacturing Automation, 2015 (10): 86-89.
[6] Carlos A. Vanegas, Daniel G. Aliaga, Bedrich B. Automatic Extraction of Manhattan-World Building Masses from 3D Laser Range Scans. IEEE Transactions on Visualization and Computer Graphic. 2012, 18(10):1627-1637.

[7] Hanyun Wang, Cheng Wang, Huan Luo, Peng Li, et al. Object Detection in Terrestrial Laser Scanning Point Clouds Based on Hough Forest[J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11:1807-1811.

[8] Dou Benjun. Research on 3D Point Cloud Data Recognition Technology of Substation Equipment Based on Laser Scanning [D]. Zhengzhou University, 2016.

[9] Li Jie, Wang Bo, Wang Xingcun, et al. 3D real-world modeling method of substation based on massive point cloud data [J]. North China Electric Power Technology, 2016 (7): 26-30.

[10] Li Lingxia. Research on classification and identification of substation equipment based on 3D point cloud [D]. Zhengzhou University, 2016.