Received 31 August 2022
Revised 26 September 2022
Accepted 26 September 2022

Robust denoising-oriented point cloud plane fitting method based on a combined model of improved Cook's distance and WTLS

Chongjun Wu, Dengdeng Shu, Hu Zhou and Zuchao Fu
College of Mechanical Engineering, Donghua University, Shanghai, China

Abstract

Purpose – In order to improve the robustness to noise in point cloud plane fitting, a combined model of improved Cook's distance (ICOOK) and WTLS is proposed by setting a modified Cook's increment, which could help adaptively remove the noise points that exceeds the threshold.

Design/methodology/approach – This paper proposes a robust point cloud plane fitting method based on ICOOK and WTLS to improve the robustness to noise in point cloud fitting. The ICOOK to denoise the initial point cloud was set and verified with experiments. In the meanwhile, weighted total least squares method (WTLS) was adopted to perform plane fitting on the denoised point cloud set to obtain the plane equation.

Findings – (a) A threshold-adaptive Cook's distance method is designed, which can automatically match a suitable threshold. (b) The ICOOK is fused with the WTLS method, and the simulation experiments and the actual fitting of the surface of the DD motor are carried out to verify the actual application. (c) The results shows that the plane fitting accuracy and unit weight variance of the algorithm in this paper are substantially enhanced.

Originality/value – The existing point cloud plane fitting methods are not robust to noise, so a robust point cloud plane fitting method based on a combined model of ICOOK and WTLS is proposed. The existing point cloud plane fitting methods are not robust to noise, so a robust point cloud plane fitting method based on a combined model of ICOOK and WTLS is proposed.

Keywords Improved cook distance, Weighted total least squares, Point cloud plane fitting, DD motor

Paper type Research paper

1. Introduction

Efficient point cloud data processing could help to improve the plane matching and modeling quality. Point cloud plane fitting is the basis of the scattered point cloud fitting algorithm, which has been a hot research topic in recent years (Xu *et al.*, 2019; Guo *et al.*, 2021; Chithra and Christoper, 2018). This technology was first used in 3D laser scanning on the ground, such as road data fitting (Narksri *et al.*, 2018; Oude Elberink and Vosselman, 2009), wall flatness detection (Wang *et al.*, 2021a; Li *et al.*, 2020), etc. Moreover, the technology is also used in industrial scenarios, such as parts flatness detection (He *et al.*, 2016; Meng *et al.*, 2016) and measurement of workpiece parts (Wang *et al.*, 2021b; Sun *et al.*, 2022; Li *et al.*, 2022), etc.

Funding: This work is supported by the Shanghai Natural Science Foundation (22ZR1402400). The authors wish to record the gratitude for the generous supports.

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval: The authors confirm that this manuscript is original and has not been published, it is also currently not under consideration for publication elsewhere.

Journal of Intelligent Manufacturing and Special Equipment
Vol. 3 No. 2, 2022
pp. 134-155
Emerald Publishing Limited
e-ISSN: 2633-660X
p-ISSN: 2633-6596
DOI 10.1108/JIMSE-08-2022-0013

However, the robust denoising method has always been a prerequisite for high-quality point cloud plane matching, which still needs extensive exploration.

In order to improve the robustness to noise, lots of researchers have conducted extensive research work. Li *et al.* (2016) proposed an eigenvalue method by comprehensively considering the errors in the three directions of X, Y and Z, and the fitting accuracy shows a substantial improvement. In addition, Pitkänen *et al.* (2019) added the accidental errors of the observation vector and coefficient matrix into the fitting model, and proposed the total least squares (TLS) method. The above traditional methods were based on the fact that point cloud observation was carried out under the condition of equal precision. However, in the actual data acquisition process, the entire observation process is absolutely at an unequal precision limited by factors such as equipment and environment difference, which could lead to different contribution weight for acquired point.

Considering the effect of equipment and environment differences, Wang *et al.* (2021c) proposed the weighted total least squares (WTLS) method. According to different weighting criteria, such as distance weighting and incident angle weighting, different weights to each point are generally assigned in the point cloud for plane fitting. The plane fitted by this method is more linear to the actual situation and can effectively improve the fitting accuracy, while the robustness to noise is not satisfying. Li *et al.* (2017) proposed the random sample consensus (RANSAC)–TLS fitting method by using RANSAC to remove rough points, and then adopting TLS to fit the model. The fitting effect of this method depended on the threshold selection of RANSAC, and each fitting result was different due to the randomness of RANSAC. Therefore, lots of researchers consider Cook's distance a good choice to eliminate noise data. Dhakal (2017) used Cook's distance to remove outliers before performing multiple regression on rice yield data in Nepal. Wang *et al.* (2018) combined the Cook's distance with the half-mean resampling method and proposed an outlier detection method based on the near-infrared spectral analysis data of soybean straw.

This paper proposes a robust point cloud plane fitting method based on improved Cook's distance (ICOOK) and WTLS to improve the robustness to noise in point cloud fitting. The ICOOK to denoise the initial point cloud was set and verified with experiments. In the meanwhile, WTLS method was adopted to perform plane fitting on the denoised point cloud set to obtain the plane equation. Finally, the point cloud plane data of the surface of the DD motor was collected, by applying the proposed method to fit the plane, verifying the applicability of our method in the real environment.

2. Denoising and fitting method

2.1 Point cloud plane processing

Traditional robust fitting methods, such as RANSAC, are a combined method for denoising and fitting, with a high efficiency but low accuracy. In this paper, denoising and fitting are separated and processed by an ICOOK and WTLS. This combined method not only has strong robustness to noise, but also fit the point cloud plane effectively in the noisy environment. The frames of the whole method are shown in Figure 1, which is divided into two parts. In the first filter part, an ICOOK to denoise the initial point cloud is designed, and the denoised result is the input of the second part. Then followed by the second part of fitting, a weighted total least square (WTLS) method is used to fit the denoised point clouds, and thus obtain the plane equation. The detailed ICOOK method and WTLS method will be described in detail in sections 2.2 and 2.3.

2.2 Improved Cook's distance(ICOOK) model

In this section, the filter part of proposed method is described. Section 2.2.1 mainly describes the original Cook's distance method and its existing problems, which is the inadaptive

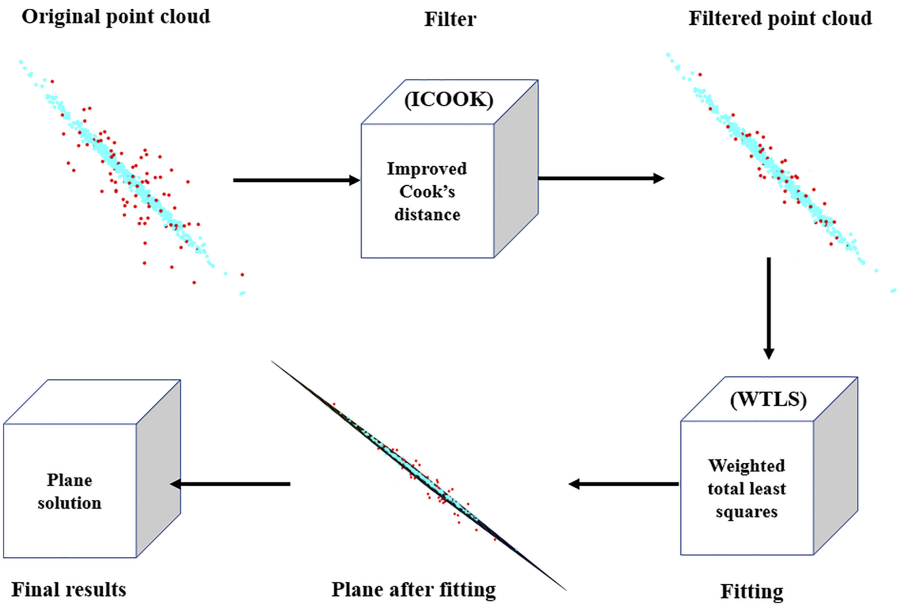


Figure 1.
Combined process of ICOOK and WTLS method

threshold selection and the low accuracy. Section 2.2.2 mainly describes the ICOOK, which is to adaptively select the Cook's distance threshold by designing an index.

2.2.1 Original Cook's distance model. In the actual situation, the collected plane point cloud will inevitably produce noise due to factors such as equipment and environment differences. How to design a criterion to identify the noise in the plane point cloud is the key process to improve the robustness of the algorithm (Peng *et al.*, 2020). The flat point cloud is assumed to be a set p, assuming that there is a threshold t, the set p can be split into two subsets: the valid point set p_1 and the noise point set p_2. The relationship is shown in formula (1).

$$\begin{aligned} p_1 \cap p_2 &= \varnothing \\ p_1 \cup p_2 &= p \end{aligned} \tag{1}$$

In this equation, the problem is transformed to find the threshold t to divide the point set p into valid point set and noise point set. In view of the fact that the plane fitting problem is a linear regression problem, the Cook's distance can be used for statistical diagnosis of the linear model, and the abnormal points can be eliminated. This paper completes the elimination of point cloud noise by calculating the Cook's distance of each point in the point set, and selecting the appropriate distance value as the threshold t. Cook's distance is a statistic parameter for statistical diagnosis of linear models. It could be given as:

$$Y = X\beta + \varepsilon \tag{2}$$

where Y is the observation vector of $n \times 1$, X is the coefficient matrix of $n \times p$, β is the parameter vector of $p \times 1$, and ε is the error vector of $n \times 1$.

The core principle of Cook's distance is to extract the i-th data from the data set when considering the influence of the i-th data on the entire model, establish a linear regression model for the remaining group data and conduct the least squares estimation (Yildiz *et al.*, 2017). Here, a quantitative function is adopted to quantify the impact of the i-th data on the entire model, which is given in formula (3).

$$D_{i,1}(M,C) = (\widehat{\beta} - \widehat{\beta}_{(i)})^T M(\widehat{\beta} - \widehat{\beta}_{(i)})/C$$
$$\widehat{\beta} = (X^T X)^{-1} X^T Y \tag{3}$$

In the formula, M is a positive definite matrix, and C is a given constant, $\widehat{\beta}$ is the initial least squares estimate and $\widehat{\beta}_{(i)}$ is the least squares estimate after extracting the i-th data. $D_{i,1}(M,C)$ measures the size of the influence of the regression coefficient estimates of β. The larger the value of $D_{i,1}(M,C)$ is, the greater the change of β after the i-th group of data is removed (Ck and Hy, 2019). Obviously, the size of $D_{i,1}(M,C)$ depends on the choice of M and C. Take $M = X^T X, C = p\widehat{\sigma}^2$ (p is the number of unknowns in the parameter vector, $\widehat{\sigma}^2$ is the unit weight variance calculated using the complete data), then the Cook's distance can be obtained:

$$D_{i,1}(X^T X, p\widehat{\sigma}^2) = \frac{(\widehat{\beta} - \widehat{\beta}_{(i)})^T X^T X(\widehat{\beta} - \widehat{\beta}_{(i)})}{p\widehat{\sigma}^2} \tag{4}$$

The larger the Cook's distance is, the greater the parameter change after excluding the first group of data would be (Hidalgo *et al.*, 2018). For the plane fitting problem in this paper, both the simulated experimental data and the measured experimental data contain some gross error values. The points with larger Cook's distance calculated in this paper are actually rough points. The larger the Cook's distance is, the more the corresponding point deviates from the fitting model is. In the meanwhile, the smaller the Cook's distance is, the closer the point is to the overall trend of the point cloud, which is a plane fitted strong influence points for plane fitting (Roberts *et al.*, 2015).

The key for plane fitting is how to set a reasonable threshold, which can not only reduce the number of misjudged noises, but also remove the real noises as much as possible. In response to this problem, two empirical criteria are adopted. The first criterion takes three times the mean value of Cook's distance as the threshold (Cook, 2000). The second criterion sets $t = 4/(n - k - 1)$, where n is the number of samples and k is the number of variables in the model to be fitted (Kim, 2017). Both of these criteria are qualitative empirical guidelines and could not be applied in some situations. To solve this problem, this paper improves the Cook's distance and designs a Cook's distance method with higher accuracy and adaptive threshold adjustment.

2.2.2 Improved cook distance model. For a linear model to be fitted, the noise points are generally less than the effective points. The Cook's distance of the valid point basically oscillates around a certain value, and only the Cook's distance values of those noise points will increase suddenly, which is similar to a wave crest. In order to apply this model in the actual situation, this paper makes the following two assumptions:

(1) In the linear model to be fitted, the proportion of noise does not exceed 30%. With the development of technology, the accuracy of point cloud acquisition equipment is getting higher and higher. It is rare that the proportion of noise in point cloud collected by equipment exceeds 30%. Therefore, the case that the noise ratio exceeds 30% is not considered.

(2) The Cook's distance values of valid points are all oscillating around a certain value. This is because that when calculating the Cook's distance of a point set by formula (4), the Cook's distance of valid points will be distributed within a small interval v, and the difference between the Cook's distance of valid points and the Cook's distance of noise points will be relatively large.

Based on the above two assumptions, only if the Cook's distance value changes drastically relative to the value of the valid point, the point is considered to be a noise point. In this model, we only need to find the point with the smallest Cook's distance and set as the threshold point. Therefore, this paper re-arranges the Cook's distance values in ascending order, which makes it easier to find the threshold point.

Figure 2 is the Cook's distance calculation before and after the ascending order of point cloud planes containing 10% noise scale. As can be seen from the figure, the Cook's distance before the ascending order is irregular. However, after sorting in ascending order, it can be seen that there is a sudden change in the Cook's distance, and the reason for the sudden

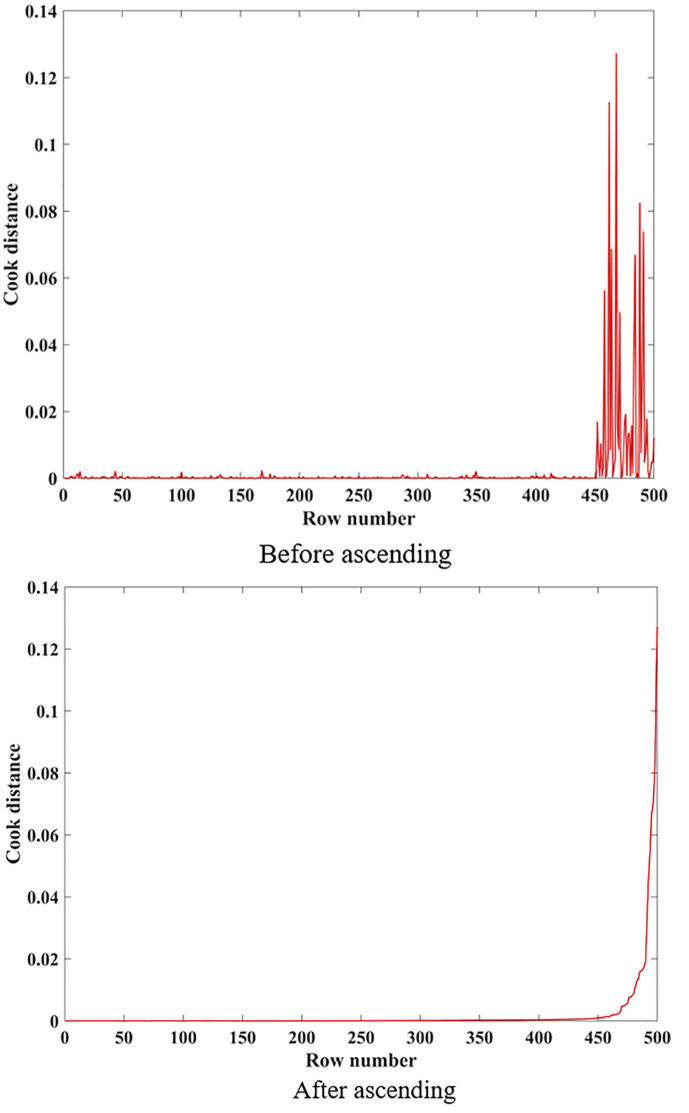


Figure 2.
Comparison of Cook's distance ascending

change is that the Cook's distance of the valid point and the Cook's distance of the noise point are quite different. Therefore, we only need to find the first point where the Cook's distance changes significantly, and set the Cook's value of this point as our threshold. In order to measure the degree of this change, an indicator as Cook's increment is designed. Thus, we need first to calculate the median value of the Cook's distance in ascending order, then subtract each Cook's distance from the median value and divide by the median value to obtain the measurement index, and finally normalize the calculated value to obtain the Cook's increment. The detailed formula is given in formula (5).

$$\theta = abs\left(\frac{cook - m}{m}\right)$$
$$norm = \frac{\theta - \min(\theta)}{\max(\theta) - \min(\theta)} \tag{5}$$

The Cook's increment actually reflects a degree of change between the Cook distance of each point and the median value. The reason why the median value is used as the reference value is that the median value is not easily affected by the extreme value, and can better represent the value that oscillates at the effective point in the hypothesis B. In order to eliminate the influence of dimensions between different data, it is also necessary to normalize the Cook's increment to the [0,1] interval. Thus, the Cook's increment can be applied to measure the degree of change of the Cook's value at each point. Through experiments in section 3.2, it is believed that the point where the Cook increment exceeds 3% as the threshold point is highly suitable.

2.3 Weighted total least squares (WTLS)

This section mainly describes the fitting part of our method. After the original data is denoised by ICOOK, a point cloud set that basically does not contain coarse noise will be obtained. Then, the WTLS method will be used to fit the denoised point cloud set.

2.3.1 WTLS fitting model. WTLS is a linear optimization method. Compared with the ordinary least squares method, this method considers both of the errors from the observation vector and the coefficient matrix. In this paper, it is assumed that the weights of each point on the model are different considering the actual situation (Wurm, 2021).

The plane equation to be fitted is as follows:

$$z_i = ax_i + by_i + c, (i = 1, 2, \ldots, n) \tag{6}$$

In this formula, a, b and c are the plane fitting parameters. Consider the coefficient matrix error and the observation vector error into the model to establish an EIV (errors-in-variable) model:

$$Z - e_z = (A - E_A) \times \xi \tag{7}$$

where Z is the $n \times 1$ observation vector with random error e_z, A is the $n \times m$ coefficient matrix with random error E_A, ξ is the parameter to be estimated. These parameters could be given as:

$$Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}, A = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{bmatrix}, \xi = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{8}$$

Moreover, the statistical properties of random errors are as follows:

$$\begin{bmatrix} e_z \\ e_A \end{bmatrix} = \begin{bmatrix} e_z \\ vec(E_A) \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_0^2 \begin{bmatrix} Q_z & 0 \\ 0 & Q_A \end{bmatrix} \right) \tag{9}$$

In this formula, *vec* means that the matrix is straightened by column, and the order is from left to right; $e_A = vec(E_A)$; σ_0^2 is the pre-test unit weight variance; $Q_{Z,}Q_A$ is the co-factor matrix of e_z and e_A, and the correlations are given as:

$$\begin{aligned} Q_Z &= {P_Z}^{-1} \\ Q_{XY} &= {P_{XY}}^{-1} \\ Q_0 &= {P_0}^{-1} \\ Q_A &= Q_0 \otimes Q_{XY} = {P_A}^{-1} \end{aligned} \tag{10}$$

In this formula, P_Z is the observation vector weight matrix, P_A is the weight matrix of the coefficient matrix A, P_0 and P_{XY} are the column vector weight matrix and the row vector weight matrix of the coefficient matrix A respectively, and $\otimes$ is the Kronecker product.

The estimation criterion for WTLS is as follows:

$$e_z^T {Q_Z}^{-1} e_z + e_A^T {Q_A}^{-1} e_A = \min \tag{11}$$

The most important step in the WTLS algorithm is the definition of the weight function. At present, there are a variety of weighting methods, such as intensity value weighting, incident angle weighting and distance weighting (Wang and Xu, 2020). Different weighting methods have different characteristics. Since the point cloud data has no intensity value and incident angle information in this paper, the distance value weighting method is adopted.

The weighted distance method assumes that the point-to-plane distance reflects the correlation between the point and the plane. The further the distance is, the lower the correlation, and the smaller the weight of the point in the fitting process, and vice versa. From this point, the weight of the point can be determined according to the distance from the point to the plane (Li *et al.*, 2015).

This paper first uses the TLS algorithm to fit an initial plane model, then calculates the distance from each point to the initial plane, and then converts the reciprocal of the distance value to a value between [0, 1] according to equation (12). In the subsequent process, the initial value of each point weight in the iterative solution process is set to be $P_{i0} = d_i'$:

$$\begin{aligned} d_i &= \sqrt{\frac{|ax_i + by_i - z_i + c|}{a^2 + b^2 + 1}} \\ d_i' &= \frac{|d_i - (\min(d_i) + \lambda)|}{\max(d_i) - \min(d_i)} \end{aligned} \tag{12}$$

In this formula, d_i is the distance from the point to the plane, $\min(d_i)$ is the minimum distance from the point to the plane, $\max(d_i)$ is the maximum distance from the point to the plane, λ is the minimum value and $\lambda = 10^{-6} mm$ in this paper.

2.3.2 The solution to the WTLS model. For the solution of the WTLS model, considering the speed and accuracy effect on the algorithm, we adopt an iterative method to solve the numerical solution of the model. The solution steps are given in Figure 3, which also represents the overall flow chart of ICOOK–WTLS method.

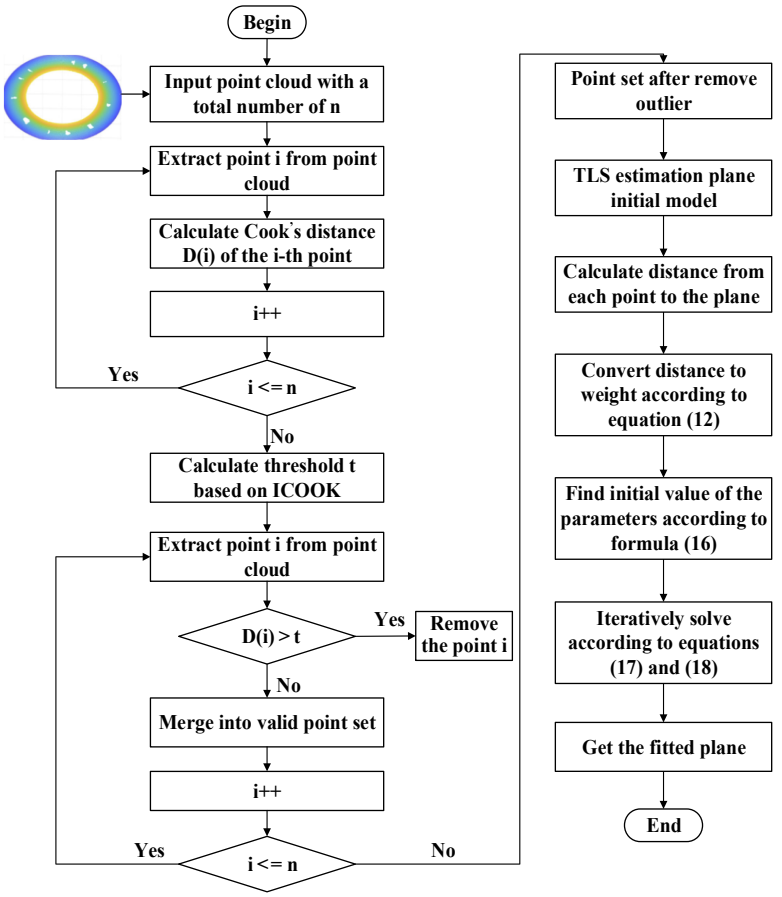


Figure 3.
Flow chart of ICOOK–WTLS

(1) In order to facilitate the iterative solution, rewriting equation (7) into the following form:

$$Z - e_z = A\xi_i + A_i\delta\xi - E_A\xi_i \tag{13}$$

In this formula, $A_i = A - \widetilde{E}_{A_i}$, $\delta\xi = \xi_{i+1} - \xi_i$, i represents the i-th iteration. Construct the Lagrangian function according to formula (14):

$$\begin{aligned}\Phi(e_z, e_A, \lambda, \xi) &= e_z^T {Q_Z}^{-1} e_z + e_A^T {Q_A}^{-1} e_A + 2\lambda^T \\ &(Z - A\xi_i - A_i\delta\xi - e_z + (\xi_i^T \otimes I_m)e_A)\end{aligned} \tag{14}$$

In this formula, λ is the $n \times 1$ Lagrangian coefficient.

(2) To solve Equation (14), the TLS method is applied to estimate the initial value of the model, and then calculate the distance from each point to the plane, convert the reciprocal of the distance value to the value of [0,1] according to Equation (12). Then, according to Equation (15), set the column vector weight matrix P_0 of A and the row

vector weight matrix $P_{XY(0)}$ of A, and the initial weight matrix $P_{Z(0)}$ of the observation value.

$$P_0 = diag[1 \quad 1 \quad 0]$$
$$P_{XY(0)} = P_{Z(0)} = diag\left[\overbrace{1 \quad 1 \quad \cdots \quad 1}^{n}\right] \tag{15}$$

(3) Solve the iterative initial value of parameter $\xi = [a, b, c]^T$ for subsequent iterations:

$$\begin{aligned} v_0 &= 0 \\ \widehat{\xi}_0 &= (A^T Q_{Z(0)} A)^{-1} A^T Q_{Z(0)} Z \\ \mu_0 &= (Q_{Z(0)} + ((\widehat{\xi}_0)^T Q_0 \widehat{\xi}_0) Q_{XY(0)})^{-1} \\ \widehat{\xi}_1 &= ((\mu_0 A)^{-1} A^T \mu_0) Z \end{aligned} \tag{16}$$

In this formula, Q_0 is the generalized inverse of P_0.

(4) Calculate μ_i, λ_i, v_i:

$$\begin{aligned} \mu_i &= (Q_{Z(i-1)} + ((\widehat{\xi}_i)^T Q_0 \widehat{\xi}_i) Q_{XY(i-1)})^{-1} \\ \lambda_i &= \mu_i (Z - A\widehat{\xi}_i) \\ v_i &= (\lambda_i)^T Q_{XY(i-1)} \lambda_i \\ \widehat{\xi}_{i+1} &= (A^T \mu_i A - v_i Q_0)^{-1} (A^T \mu_i Z) \end{aligned} \tag{17}$$

(5) According to the value calculated in step (4), calculate $\widehat{\xi}_{i+1}$:

$$\widehat{\xi}_{i+1} = (A^T \mu_i A - v_i Q_0)^{-1} (A^T \mu_i Z) \tag{18}$$

(6) Repeat steps (4) to (5) until $\|\widehat{\xi}_{i+1} - \widehat{\xi}_i\| < \delta$ and δ are given threshold.

The above process is the full theoretical analysis of our method. Our method incorporates the ICOOK and WTLS, which corresponds the denoising and fitting process.

3. Simulation and discussion

3.1 Simulation experiment

In order to verify the accuracy and robustness of the method under different noise ratios, this paper conducted extensive experiments through MATLAB2017b (64bit) platform on a personal computer, with Intel(R) Core (TM) i5-5257U CPU @ 2.70GHz. In this simulation, plane data containing 500 points are randomly generated, the plane equation is $Z = -0.5 * X + 0.6 * Y + 0.7$, X and Y are random numbers in the $[-1,1]$ interval, and a normal distribution random error with mean 0 and standard deviation 0.05 is added. The reason for choosing 0.05 for the standard deviation is that the error for the simulation is an inevitable systematic error in the experiment, and this type of error is generally small. The coordinate range of the simulated data is between $[-1, 1]$, so the added normal error cannot be too large, thus the standard deviation is set to 0.05. In order to verify the superiority of Cook's distance in denoising with different noise ratios, 5%, 10%, 15% and 30% gross error data

were added to the plane data, respectively. The gross data follow a normal distribution with a mean of 0 and a standard deviation of 0.5. In order to facilitate verification, the number of noise points added is at the end. For example, in data with 5% noise, there are 25 noise points, and the point number is within 475–500.

The simulation results are depicted in Figure 4, which is mainly focused on the performance of the proposed method from two aspects. This is first to verify that our method on adaptive performance of the Cook's distance threshold and the corresponding denoising performance. Moreover, another advantage is to verify that the point cloud plane fitted by our method is more accurate and robust. It could be seen from Figure 4 that the point cloud is more scattered with the increase of noise ratio.

3.2 Evaluating denoising results of ICOOK

The threshold points are obtained by applying the ICOOK algorithm in the data of the above four different proportions of noise, and the results are shown in Figure 5. Figure 5 is the calculated Cook's increment graph with a threshold line of 3%, which has been discussed in section 2.3. The threshold points of the four different noise ratios are the intersection points of the threshold line and the polyline. The corresponding Cook's distance value is obtained by the serial number of the intersection point, and this value is the final Cook's distance

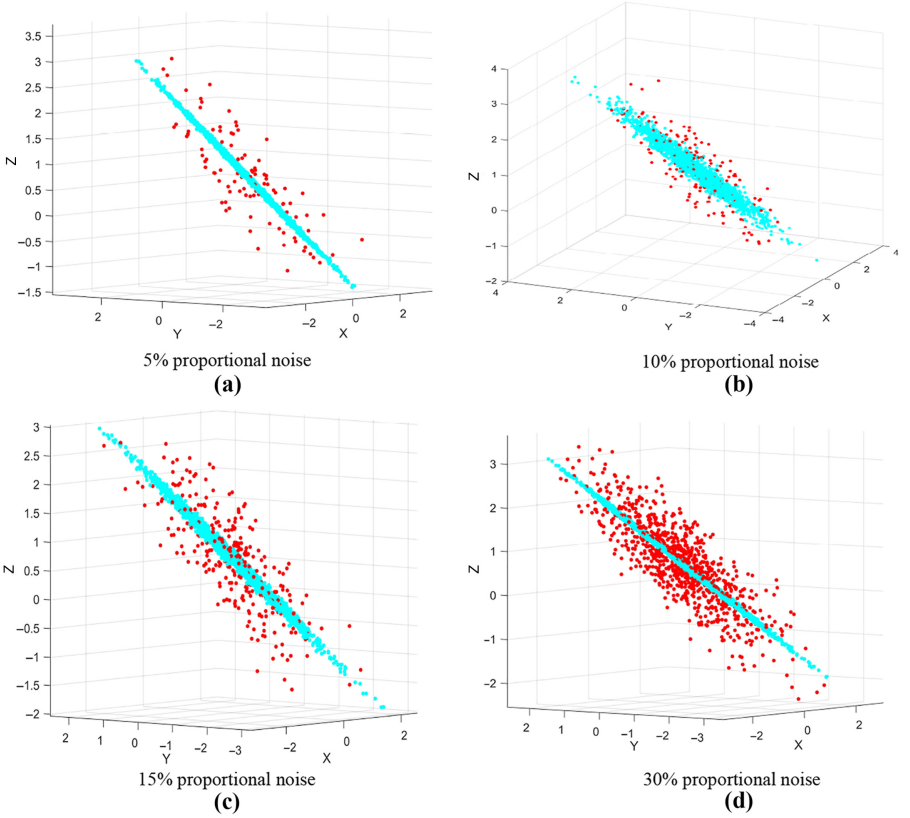


Figure 4.
Simulated point cloud planes with different noise scales

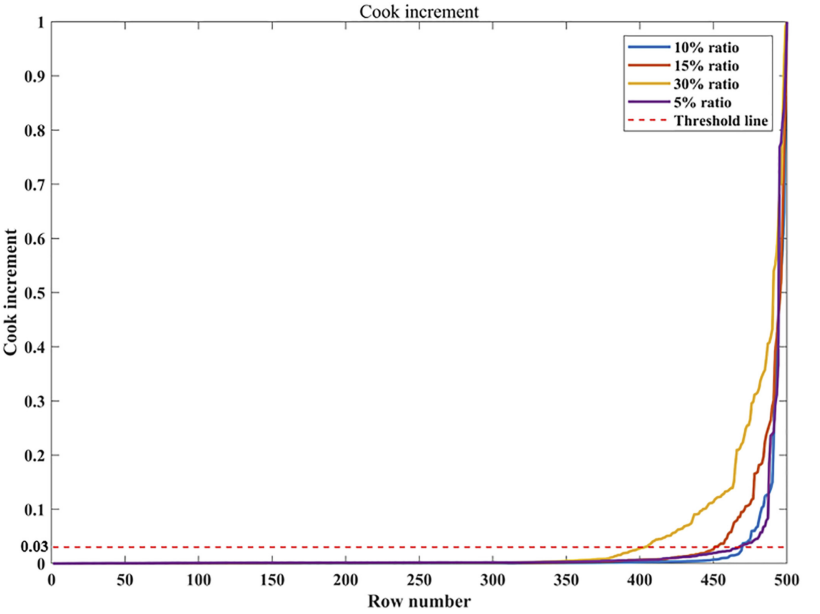


Figure 5.
Cook's increment

threshold. The thresholds obtained by the ICOOK algorithm and the thresholds of the two empirical criteria are shown in Table 1.

In order to achieve a better threshold, all three thresholds are plotted on the corresponding Cook's distance map, and the results are shown in Figure 6. In this figure, the abscissa is the serial number of the point, the ordinate is the Cook's distance value, the blue line is the standard 2 threshold, that is $t = 4/(n - k - 1)$, the green line is the standard 1 threshold, that is three times the mean value of the Cook's distance, and the red line is the threshold in this paper. Among the four noise ratios, the threshold in this paper is the smallest, and the threshold of standard 2 is the highest. It can be seen that the threshold of standard 2 is set too high, resulting in the most missed detection noise. While the threshold of standard 1 is second, which is especially obvious when the noise ratio increases, the threshold value of ICOOK changes adaptively with the change of the noise ratio, and the missed detection noise is the least among the three threshold values.

In order to quantitatively evaluate the denoising effect, the noise points are counted in this paper. In terms of the denoising algorithms, two indicators are generally considered. First, which kind of noise type is filtered. Second, how much non-noise is filtered as noise, which is a misjudgment rate. This is essentially a classification evaluation problem, and this paper draws on two classic evaluation metrics in machine learning: recall and precision

Table 1.
Different noise ratio threshold

Noise ratio (%)	Threshold point number	Corresponding threshold	Standard 1 threshold	Standard 2 threshold
5	468	0.0028	0.0048	0.0080
10	470	0.0048	0.0063	0.0080
15	451	0.0025	0.0054	0.0080
30	404	0.0015	0.0060	0.0080

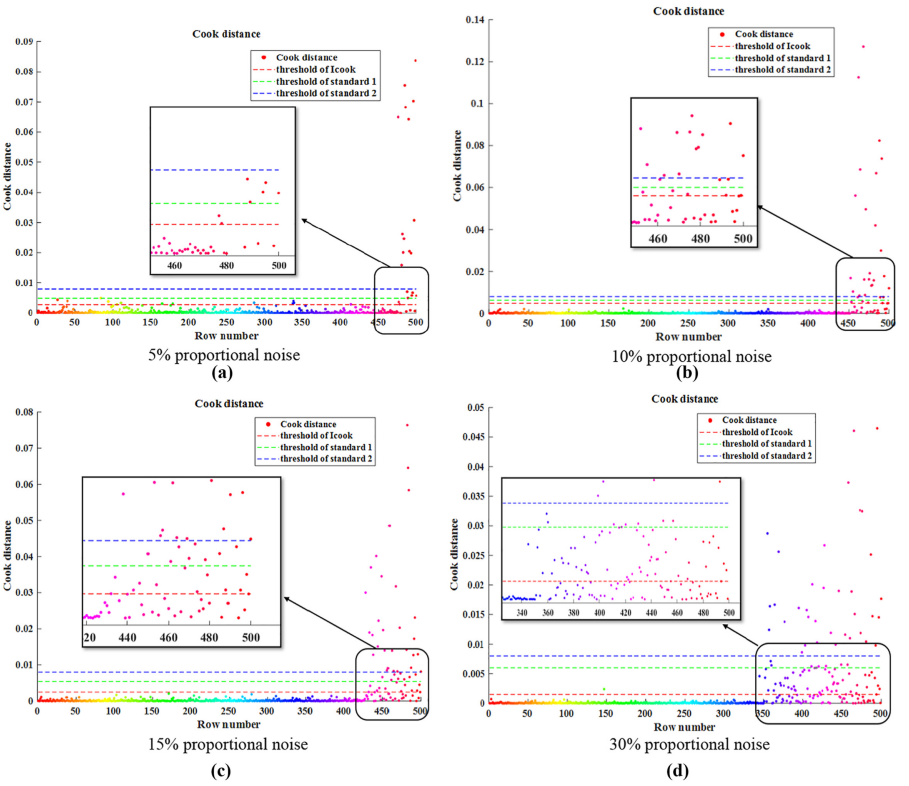


Figure 6.
Effect of different thresholds on Cook's distance

(Berkel *et al.*, 2018). The recall rate reflects the degree of true noise filtering, and the precision rate reflects the degree of false detection by the algorithm. The definitions of the two indicators are shown in Equation (19):

$$P = \frac{TP}{TP + FP}$$
$$R = \frac{TP}{TP + FN} \tag{19}$$

where P and R represents the precision rate and the recall rate. TP is the detection noise, FP is the false detection noise, and FN is the missed detection noise. The F1 metric (Chicco and Jurman, 2020) is introduced based on the influence of the recall rate and precision rate, and the formula is as follows.

$$F1 = \frac{2 \times P \times R}{P + R} \tag{20}$$

The F1 measurement can synthesize the influence of the recall rate and the precision rate. In the performance comparison, only the F1 indicator needs to be compared. The larger the F1 is, the better the effect. The statistical results are shown in Table 2. In order to compare the

Threshold	Noise ratio (%)	Detected noise point	False detection noise pint	Missed detection noise point	Non-noise point	Recall	Precision	F1
Threshold of standard 1	5	18	1	7	474	0.72	0.95	0.82
	10	30	0	20	450	0.6	1	0.75
	15	52	0	23	425	0.69	1	0.82
	30	73	0	77	350	0.49	1	0.66
Threshold of standard 2	5	13	0	12	475	0.52	1	0.68
	10	34	0	16	450	0.68	1	0.81
	15	40	0	35	425	0.53	1	0.69
	30	61	0	89	350	0.41	1	0.58
Threshold of ICOOK	5	23	4	5	468	0.82	0.85	0.83
	10	40	0	10	450	0.80	1	0.89
	15	63	2	12	423	0.84	0.97	0.90
	30	108	3	42	347	0.72	0.97	0.83

Table 2.
Comparison of denoising results with different thresholds

relationship between the data more clearly, the F1 index values in Table 2 are drawn into a three-dimensional histogram, as shown in Figure 7.

As can be seen in Figure 7, the F1 value obtained from the threshold value of standard 2 generally shows a decreasing trend with the increase of the noise ratio, jumping between 0.6 and 0.8, and the fluctuation is large. The F1 value calculated by the threshold of standard 1 does not change much with the increase of the noise ratio, and the overall stability is around 0.7. The reason for this is that the threshold of standard 2 is only related to the number of samples and variables, and has nothing to do with the noise ratio. In this paper, the threshold of standard 2 is a fixed value. With the increase of the noise ratio, the Cook's value is changing, but the corresponding threshold does not change, so the noise that is missed is more and more. And the threshold of standard 1 is related to the mean value of Cook's distance. As the noise ratio increases, the mean value of Cook's distance will change, and the corresponding threshold is also adjusted adaptively. Therefore, the result of standard 1 is not

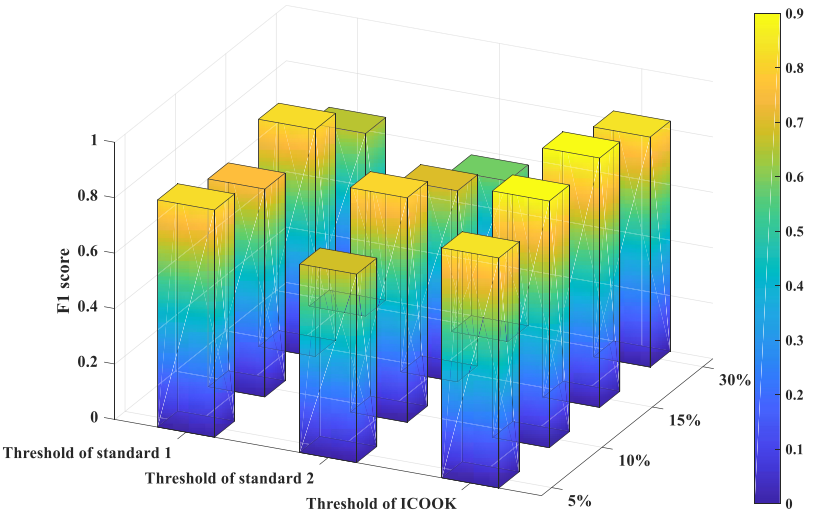


Figure 7.
Three dimensional histogram of denoising results with different thresholds

sensitive to the change of noise adaptively. In contrast to the algorithm in this paper, it can be found that the F1 value obtained by the algorithm in this paper does not fluctuate much, and is basically stable at 0.86. It is not sensitive to the change of noise ratio and has strong adaptability. Besides, in all noise ratios, the highest F1 value is the algorithm in this paper, which shows that the algorithm in this paper is more adaptive and accurate.

3.3 Evaluating plane fitting results of WTLS

In Section 3.2, the denoising effect of the ICOOK algorithm is evaluated. Compared with the other two threshold criteria, the denoising accuracy and robustness of the ICOOK method are much better. In this section, the point cloud plane will be fitted. In order to evaluate the accuracy after fitting, the LS and WTLS algorithms are adopted to compare with the algorithm in this paper.

In the experiment, the fitting quality of the algorithm is evaluated by the two indicators of plane fitting accuracy and unit weight variance. Among them, the plane fitting accuracy is essentially the mean square error of the model. The smaller the fitting accuracy is, the higher the fitting accuracy of the algorithm. This metric can evaluate the accuracy of the estimation results (Ou *et al.*, 2020). Another indicator, the unit weight variance, can directly analyze the difference between the fitting methods, and can better reflect the performance of the estimation method itself. The smaller the variance in the unit weight, the better the fitting performance (Gong and Li, 2014). The definitions of the two indicators are shown in equations (20) and (21).

$$\sigma_p = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} \tag{21}$$

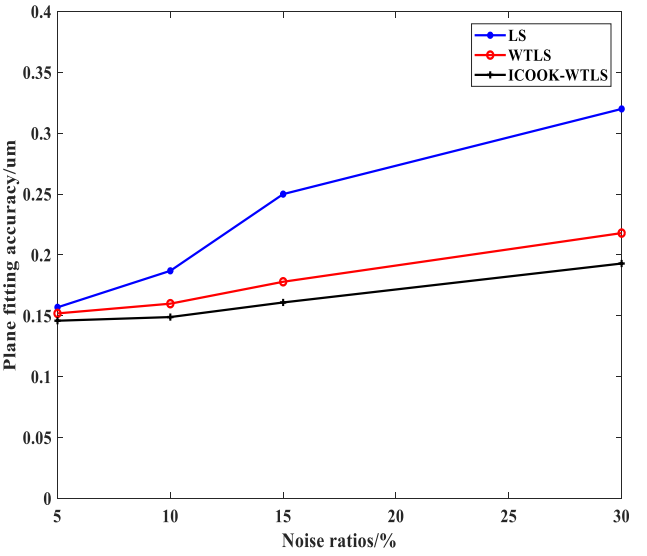
$$\sigma_0 = \sqrt{\frac{\lambda^T (Z - A\xi)}{n - 3}} \tag{22}$$

In the formula, σ_p is the plane fitting accuracy, σ_0 is unit weight variance, n is the total number of point clouds and di is the distance from the point to the plane in formula (12).

The above point cloud data are fitted with LS, WTLS and ICOOK–WTLS (algorithm in this paper) respectively. The calculated plane parameters and the accuracy evaluation index are shown in Table 3, and the accuracy analysis diagram drawn from the analysis of the data in the table is shown in Figure 8. It can be seen that when the gross error ratio is 5%, the obtained coefficient value is very close to the actual value. The fitting accuracy

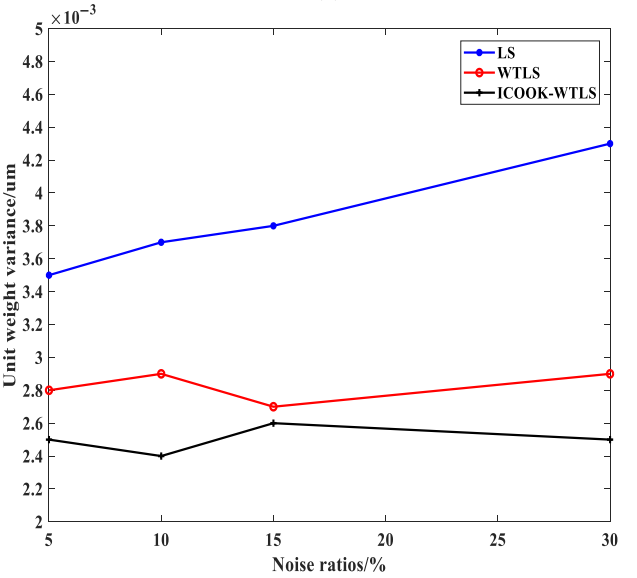
Table 3. Data comparison of simulation algorithms

Noise ratio (%)	Algorithm	a(–0.5)	b(0.6)	c(0.7)	$\sigma_0/\mu m$	$\sigma_p/\mu m$
5	LS	−0.498	0.595	0.702	0.0035	0.157
	WTLS	−0.498	0.596	0.703	0.0028	0.152
	ICOOK–WTLS	−0.500	0.595	0.703	0.0025	0.146
10	LS	−0.511	0.609	0.697	0.0037	0.187
	WTLS	−0.502	0.607	0.695	0.0029	0.160
	ICOOK–WTLS	−0.502	0.605	0.696	0.0024	0.149
15	LS	−0.475	0.587	0.708	0.0038	0.25
	WTLS	−0.491	0.594	0.705	0.0027	0.178
	ICOOK–WTLS	−0.492	0.596	0.705	0.0026	0.161
30	LS	−0.518	0.586	0.715	0.0043	0.32
	WTLS	−0.500	0.595	0.708	0.0029	0.218
	ICOOK–WTLS	−0.499	0.592	0.709	0.0025	0.193



Plane fitting accuracy comparison under different noise ratios

(a)



Unit weight variance comparison under different noise ratios

(b)

Figure 8.
Algorithm accuracy in simulation experiments

and unit weight variance of each algorithm are not much different, the maximum difference of fitting precision is only 0.011, and the maximum difference of unit weight variance is 0.001, which shows that the fitting effect of each method is good. Only due to a small amount of noise, the fitting effect of each method has a small deviation. When the gross error ratio gradually increases from 10% to 30%, it can be seen from Figure 8 that both the fitting accuracy and the unit weight variance show an increasing trend, which shows that the fitting effect of each algorithm deteriorates with the increase of the gross error ratio.

Among them, with the increase of the gross error ratio, the fitting effect of the LS algorithm decreases most obviously, indicating that the LS algorithm has poor robustness to gross error data; WTLS uses the distance of each point relative to the initial fitting plane as a weight matrix. The farther the distance is, the smaller the weight, which can reduce the influence of noise to a certain extent. However, when the proportion of gross error increases gradually, the difference between the two indicators of the plane fitting accuracy and the unit weight variance of the WTLS algorithm is getting bigger, and the fitting effect is getting worse. This shows that the robustness of WTLS to noise is not strong enough.

On the basis of WTLS, the algorithm in this paper combines Cook's distance to remove gross error values, and integrates the superior denoising performance of Cook's distance into WTLS. It can be seen from Table 3 and Figure 8 that ICOOK–WTLS has the best indicators among the three algorithms, and with the increasing proportion of gross errors, compared to LS and WTLS, ICOOK–WTLS algorithm can still guarantee a good fitting effect, and the robustness to noise is greatly enhanced.

4. Experiment verification

Simulation experiments show that our method maintains good performance under different noise ratios. To further verify the applicability of this method, the laser sensor is used to collect the point cloud of the DD motor surface and the surface of the part with holes for experimental verification. Since the DD motor surface and the surface of the part with holes is generally applied in high-precision positioning, their flatness accuracy is very high. Hikvision's MV-DP2305-01H 3D laser scanner is used to collect the point cloud. The scanner is a micron scale scanner and the relevant parameters are shown in Table 4.

In order to verify the obtained plane equation, the scanner is placed horizontally, and the DD motor and the part with holes are placed horizontally to the scanner. The heights of their surfaces from the scanner are 6130 and 6345μm, respectively. From this, the plane equation of the surface of the DD motor and the part with holes can be obtained as $Z = 6{,}130$ and $Z = 6{,}345$.

The experimental platform and scanned point cloud data are shown in Figure 9. The experiment was carried out in an indoor environment with uniform illumination without external light source. There are 3 main sources of the noise in the point cloud of DD motor. First, there are many screw holes in the surface of the DD motor for installing parts. During the acquisition process, when the laser is irradiated into the holes, the reflectivity will be greatly reduced, which directly leads to many noise spots around the holes. These noises are probably 20% of the total point cloud. Then, the edge of the scanned point cloud data also has noise, which accounts for about 5%. Finally, due to uncontrollable factors such as the

Table 4. Parameters of linear laser sensor

Points of single contour	Near field	Far field	Measuring range	*Z*-axis resolution	*Z*-axis repeatability	Scanning frequency
2048	25.2 mm	34.4 mm	25 mm	1.8~3.0 μm	0.4 μm	700 Hz

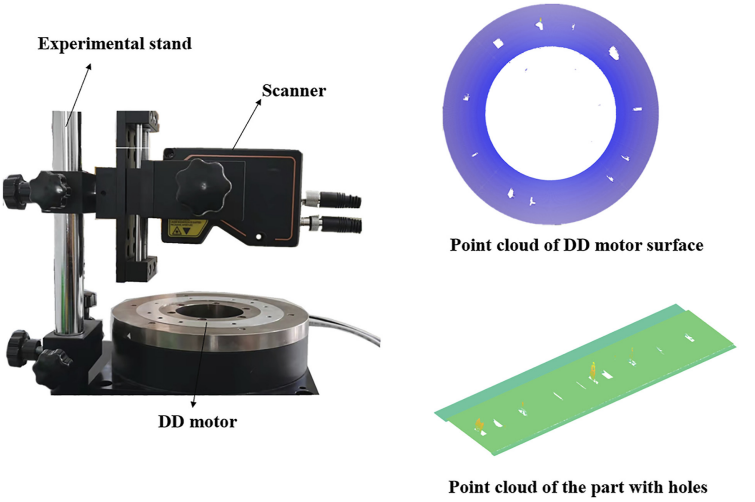


Figure 9.
Experimental platform and point cloud

environment and equipment, noise can also be generated, and this kind of noise accounts for about 5% of the total point cloud.

In the point cloud of the part with holes, there are two main sources of noise. The uncorrelated facets could make up 10% of the total point cloud. Moreover, the noise around the hole accounts for 5% of the total point cloud. In order to correspond to the four noise ratios set in the simulation experiment, the point cloud with 5 and 30% noise ratio is segmented from the surface point cloud of the DD motor, and the point cloud with 10 and 15% noise ratio is segmented from the point cloud of parts with holes. The results are shown in Figure 10.

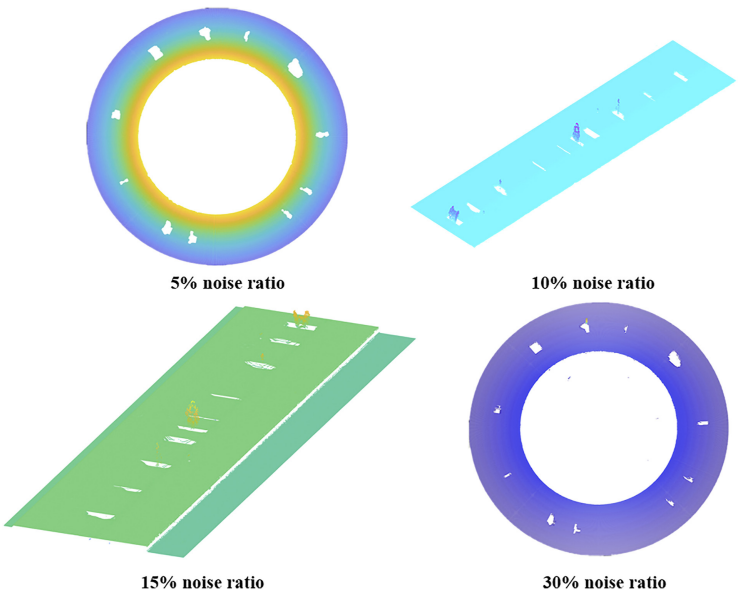


Figure 10.
Point cloud with different noise ratio

Four algorithms are used to fit the above point cloud. The fitting results are shown in Table 5. As can be seen from Table 5, the normal vectors of the plane equations solved by the three methods under the two noise ratios are all (0, 0, 1), and the difference is mainly reflected in the different fitting heights. The error between the theoretical value and the fitted value is plotted as a line graph, as shown in Figure 11.

As can be seen from Figure 11 that the error of the LS algorithm is the highest under the four noise ratios, which shows that the LS algorithm is not robust to noise. In the case of noise, the effect of the LS algorithm is relatively poor. The WTLS algorithm and the RANSAC algorithm are robust fitting algorithms, which are robust to noise. In the presence of noise, the error values of the two algorithms can be kept at a relatively low level. However, when the

Noise ratio (%)	Algorithm	a	b	c	Error	$\sigma_O/\mu m$	$\sigma_p/\mu m$
5	LS	4.647e−6	3.967e−6	6121.64	1.36e−3	0.151	18.541
	WTLS	5.156e−6	3.327e−6	6132.13	3.47e−4	0.142	17.956
	RANSAC	5.079e−6	3.087e−6	6131.24	2.02e−4	0.138	17.916
	ICOOK–WTLS	4.786e−6	2.658e−6	6130.34	5.55e−5	0.125	17.892
10	LS	3.148e−6	6.757e−6	6354.52	1.50e−3	0.159	18.967
	WTLS	3.559e−6	6.047e−6	6346.11	1.75e−4	0.149	18.023
	RANSAC	3.416e−6	5.921e−6	6345.94	1.48e−4	0.144	18.386
	ICOOK–WTLS	3.466e−6	5.456e−6	6345.53	8.35e−5	0.128	17.916
15	LS	4.647e−6	7.127e−6	6357.21	1.92e−3	0.165	19.248
	WTLS	4.976e−6	6.925e−6	6346.56	2.46e−4	0.156	18.169
	RANSAC	4.906e−6	6.529e−6	6347.15	3.39e−4	0.163	19.159
	ICOOK–WTLS	4.647e−6	6.018e−6	6345.67	1.06e−4	0.134	18.012
30	LS	7.387e−6	5.256e−6	6143.32	2.17e−3	0.173	19.765
	WTLS	4.257e−6	4.586e−6	6134.78	7.80e−4	0.151	18.236
	RANSAC	4.946e−6	4.829e−6	6133.51	5.73e−4	0.155	19.516
	ICOOK–WTLS	5.134e−6	3.367e−6	6132.57	4.19e−4	0.138	18.132

Table 5.
Data comparison of different algorithms in practical experiments

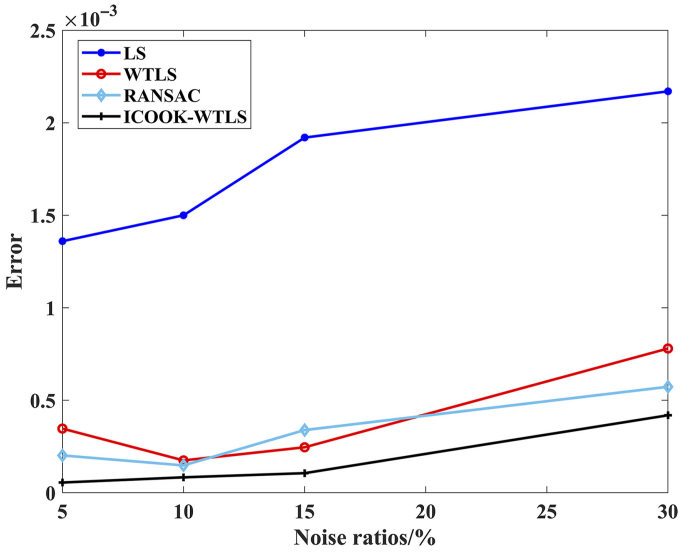


Figure 11.
Error picture

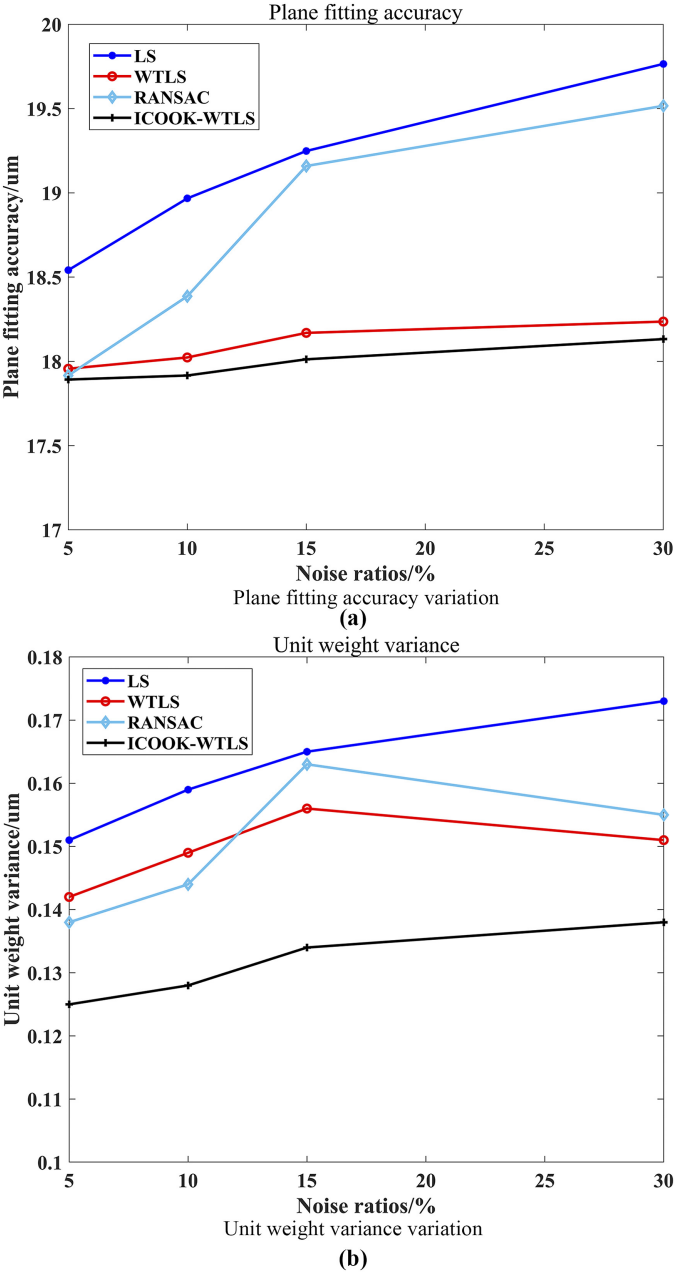


Figure 12.
Algorithm accuracy in real experiments

noise ratio exceeds 15%, the error values of the two algorithms increase significantly, and the fitting effect becomes worse. This shows that the WTLS algorithm and the RANSAC algorithm have a certain robustness to noise, but when the noise ratio exceeds 15%, the robustness will decrease significantly. In the algorithm of this paper, the error value is the

smallest under the four noise ratios. Even if the noise ratio exceeds 15%, the error value of the algorithm in this paper is still kept in a small range, and it is still very robust to noise.

The plane fitting precision and unit weight variance can reflect the fitting precision of the algorithm. The lower the two values, the higher the fitting precision of the algorithm. Line graphs of plane fit accuracy and unit weight variance are plotted in Figure 12. As can be seen from Figure 12, with the increase of the noise ratio, the performance of the four methods decreases, and the LS method is most affected by the noise change. When the noise ratio increases to 30%, the two index values of the LS method increase sharply, and the algorithm in this paper has the smallest change. This shows that in the measured data, the algorithm in this paper has the highest accuracy, the LS algorithm performs the worst.

The accuracy of WTLS algorithm is higher than LS algorithm, and it is not much different from the accuracy of RANSAC algorithm. The accuracy of the RANSAC algorithm fluctuates greatly, because the RANSAC algorithm needs to set different thresholds for different proportions of noise, and the algorithm has a certain randomness, which causes the algorithm accuracy to fluctuate greatly. The accuracy of this algorithm is the highest. Because the algorithm incorporates the ICOOK, it can adaptively set the denoising threshold and remove abnormal data from the point cloud, which improves the robustness of the algorithm to noise. To sum up, under the four different noise ratios, the algorithm in this paper not only has the smallest fitting error and highest fitting accuracy with a strong robustness to noise.

5. Conclusion

In order to improve the robustness to noise in point cloud plane fitting, a combined model of ICOOK and WTLS is proposed by setting a modified Cook's increment, which could help adaptively remove the noise points that exceeds the threshold. The main contributions of the algorithm could be summarized as:

(1) A threshold-adaptive Cook's distance method is designed, which can automatically match a suitable threshold. Compared with the current adopted threshold, this method shows a better denoising performance and adaptability in simulated point cloud sets under noise ratios from 5% to 30%.

(2) The ICOOK is fused with the WTLS method, and the simulation experiments and the actual fitting of the surface of the DD motor are carried out to verify the actual application.

(3) The results shows that the plane fitting accuracy and unit weight variance of the algorithm in this paper are substantially enhanced. This shows that the algorithm in this paper is robust enough to noise, and effectively solves the problem that the current point cloud plane fitting method is not robust enough to noise.

References

Berkel, N.V., Goncalves, J., Lovén, L., Ferreica, D., Hosio, S. and Kostakos, V. (2018), "Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports", *International Journal of Human-Computer Studies*, Vol. 125, pp. 118-128.

Chicco, D. and Jurman, G. (2020), "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation", *BMC Genomics*, Vol. 21 No. 1, pp. 1-13.

Chithra, P. and Christoper, T. (2018), "3D color point Cloud compression with plane fitting and discrete wavelet transform", *Tenth International Conference on Advanced Computing*, (ICoAC), pp. 20-26.

Ck, A. and Hy, B. (2019), "Diagnostic measures for kernel ridge regression on reproducing kernel Hilbert space", *Journal of the Korean Statistical Society*, Vol. 48 No. 3, pp. 454-462.

Cook, R.D. (2000), "Detection of influential observation in linear regression", *Technometrics*, Vol. 42 No. 1, pp. 65-68.

Dhakal, C.P. (2017), "Dealing with outliers and influential points while fitting regression", *Journal of Institute of Science and Technology*, Vol. 22 No. 1, pp. 61-65.

Gong, X. and Li, Z. (2014), "A robust weighted total least squares method", *Acta Geodaetica et Cartographica Sinica*, No. 9, pp. 888-894.

Guo, W., Wu, C., Ding, Z. and Zhou, Q. (2021), "Prediction of surface roughness based on a hybrid feature selection method and long short-term memory network in grinding", *The International Journal of Advanced Manufacturing Technology*, Vol. 112, pp. 2853-2871.

He, W., Cao, X., Kuang, L., Zhang, P., Zhang, Y., Wu, F., Xu, M. and To, S. (2016), "Detection for flatness of large surface based on structured light", *Eighth International Symposium on Advanced Optical Manufacturing and Testing Technology*, Vol. 9684, doi: 10.1117/12.2243409.

Hidalgo, J.A.P., Pérez-Suay, A., Nar, F., Camps-Valls, G. (2018), "Nonlinear Cook distance for anomalous change detection", *IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium,* IEEE, pp. 5025-5028.

Kim, M.G. (2017), "A cautionary note on the use of cook distance", *Communications for Statistical Applications and Methods*, Vol. 24 No. 3, pp. 317-324.

Li, M.F., Ou, J.X., Tan, D. and Yang, H.H. (2015), "Study on fixed weight methods in plane fitting of point clouds based on weighted total least squares", *Journal of Geodesy and Geodynamics*, Vol. 35 No. 3, pp. 428-432.

Li, Y., Geng, G. and Jiang, Q. (2016), "A parallel contour integral method for eigenvalue analysis of power systems", *IEEE Transactions on Power Systems*, pp. 624-632.

Li, X., Han, K. and Xiong, F.G. (2017), "Plane fitting of point clouds based on RANSAC and TLS", *Computer Engineering and Design*, Vol. 38 No. 1, p. 5.

Li, D., Liu, J., Feng, L., Zhou, Y., Liu, P.K. and Chen, Y.F. (2020), "Terrestrial laser scanning assisted flatness quality assessment for two different types of concrete surfaces", *Measurement*, Vol. 154, 107436.

Li, C., Piao, Y., Meng, B., Hu, Y.X., Li, L.Q. and Zhang, F.H. (2022), "Phase transition and plastic deformation mechanisms induced by self-rotating grinding of GaN single crystals", *International Journal of Machine Tools and Manufacture*, Vol. 172, 103827.

Meng, J., Qi, L. and Liang, Y. (2016), "Flatness detection research based on the string vibration method", *12th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, IEEE, pp. 1587-1592.

Narksri, P., Takeuchi, E., Ninomiya, Y., Morales, Y., Akai, N., Kawaguchi, N. (2018), "A slope-robust cascaded ground segmentation in 3D point cloud for autonomous vehicles", *2018 21st International Conference on Intelligent Transportation Systems (ITSC),* IEEE, pp. 497-504.

Ou, J.X., Deng, X.W., Cai, M.X. and Qiu, M. (2020), "A robust methods of fitting plane based on LS-WTLS", *Laser Technology*, Vol. 44 No. 6, pp. 784-788.

Oude Elberink, S. and Vosselman, G. (2009), "3D information extraction from laser point clouds covering complex road junctions", *The Photogrammetric Record*, Vol. 24 No. 125, pp. 23-36.

Peng, F., Yu, H., Tao, J., Su, Y.F., Zhou, D., Zeng, X. and Li, X. (2020), "Efficient statistical analysis for correlated rare failure events via asymptotic probability approximation", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, No. 99, p. 1.

Pitkänen, T.P., Raumonen, P. and Kangas, A. (2019), "Measuring stem diameters with TLS in boreal forests by complementary fitting procedure", *Journal of Photogrammetry and Remote Sensing*, Vol. 147 January, pp. 294-306.

Roberts, S., Martin, M.A. and Zheng, L. (2015), "An adaptive, automatic multiple-case deletion technique for detecting influence in regression", *Technometrics*, Vol. 57 No. 3, pp. 408-417.

Sun, Y., Jin, L., Gong, Y., Wen, X.L., Yin, G.Q., Wen, Q. and Tang, B.J. (2022), "Experimental evaluation of surface generation and force time-varying characteristics of curvilinear grooved micro end mills fabricated by EDM", *Journal of Manufacturing Processes*, Vol. 73, pp. 799-814.

Wang, L.Y. and Xu, R.R. (2020), "Robust weighted total least squares algorithm for three-dimensional coordinate transformation", *Journal of Geodesy and Geodynamics*, Vol. 40 No. 10, p. 7.

Wang, Y., Liu, Q., Hou, H., Rho, S.M., Gupta, B., Mu, Y.X. and Shen, W.Z. (2018), "Big data driven outlier detection for soybean straw near infrared spectroscopy", *Journal of Computational Science*, Vol. 26, pp. 178-189.

Wang, Q., Yang, R., Wu, C. and Liu, Y. (2021), "An effective defect detection method based on improved Generative Adversarial Networks (iGAN) for machined surfaces", *Journal of Manufacturing Processes*, Vol. 65, pp. 373-381.

Wang, J., Yan, Y., Li, Z. and Geng, Y.Q. (2021), "Towards understanding the machining mechanism of the atomic force microscopy tip-based nanomilling process", *International Journal of Machine Tools and Manufacture*, Vol. 162, 103701.

Wang, J., Zhao, J., Liu, Z. and Kang, Z.J. (2021), "Location and estimation of multiple outliers in weighted total least squares", *Measurement*, Vol. 181, 109591.

Wurm, M. (2021), "A universal and fast method to solve linear systems with correlated coefficients using weighted total least squares", *Measurement Science and Technology*, Vol. 33 No. 1, 015017.

Xu, D., Li, F. and Wei, H. (2019), "3D point cloud plane segmentation method based on RANSAC and support vector machine", *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, IEEE, pp. 943-948.

Yildiz, B., Bilbao, J.I. and Sproul, A.B. (2017), "A review and analysis of regression and machine learning models on commercial building electricity load forecasting", *Renewable and Sustainable Energy Reviews*, Vol. 73, pp. 1104-1122.

Corresponding author
Chongjun Wu can be contacted at: wcjunm@dhu.edu.cn