

---

# Surface reconstruction method for measurement data with outlier detection by using improved RANSAC and correction parameter

Tianqi Gu<sup>1</sup>, Zude Luo<sup>1</sup>, Dawei Tang<sup>2</sup>, Jianxiong Chen<sup>1</sup>, Shuwen Lin<sup>1</sup>

<sup>1</sup>School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China

<sup>2</sup>Centre for Precision Technologies, University of Huddersfield, Huddersfield HD1 3DH, UK

**Corresponding author:**

Shuwen Lin, School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China

E-mail: swl2010@126.com

**Abstract:** The moving least squares (MLS) and moving total least squares (MTLS) are two of the most popular methods used for reconstructing measurement data, on account of their good local approximation accuracy. However, their reconstruction accuracy and robustness will be greatly reduced when there are outliers in measurement data. This article proposes an improved MTLS method (IMTLS), which introduces an improved random sample consensus (RANSAC) algorithm and a correction parameter in the support domain, to deal with the outliers and random errors. Based on the nodes within the support domain, firstly the improved RANSAC is used to generate a model for establishing the group of pre-interpolation and calculating the residual of each node. Subsequently, the abnormal degree of the node with the largest residual is evaluated by the correction parameter associated with the node residual and random errors. The node with certain abnormal degree will be eliminated and the remaining nodes are used to obtain the approximation coefficients. The correction parameter can be used for data reconstruction without insufficient or excessive elimination. The results of numerical simulation and measurement experiment show that the reconstruction accuracy and robustness of the IMTLS method is superior to the MLS and MTLS method.

**Keywords:** Measurement data, Outliers, Moving total least squares, Random sample consensus, Surface reconstruction

## 1. Introduction

After decades of development, the reconstruction algorithms for discrete data have played an essential role in many engineering and scientific fields, especially in error analysis and data processing. The conventional numerical methods, such as finite element method (FEM), interpolate or approximate the nodes through defining a mesh based on known nodes.<sup>1</sup> However, the accuracy of reconstruction will be greatly reduced and the fitting will even fail when the grid-based method is used to deal with large deformation and discontinuity problems.<sup>2</sup> In addition, the human-labor and time cost of generating meshes in complex-shaped domains are not satisfactory. While meshless methods use node-based approximation without mesh discretization, the efficiency and accuracy of data processing are, therefore, greatly improved.<sup>3</sup> In order to meet the development requirement of different fields, a multitude of meshless methods have been presented and applied, such as the

element-free galerkin (EFG),<sup>4</sup> diffuse elements method (DEM)<sup>5</sup> and moving least squares (MLS).

In 1974, Mclain<sup>6</sup> proposed the weighted least squares method. On this basis, Lancaster and Salkauskas<sup>7</sup> introduced the moving concept and proposed the MLS method in 1981. To this day, the MLS has become an important method for constructing shape function. Unlike the traditional least square method using complete polynomials, the shape function of the MLS is composed of a coefficient vector and a basis function vector, which can obtain higher continuity under the condition with low order basis function.<sup>8</sup> The introduction of the weight function with compact support makes the reconstructed curve or surface accurate and smooth, which has contributed to the wide application of the MLS in various fields. For example, the MLS is applied to solve elasticity problems,<sup>9</sup> the compressible Navier-Stokes,<sup>10</sup> Kuramoto-Sivashinsky<sup>11</sup> and Burgers equation,<sup>12</sup> and estimate mathematical model based on discrete points.<sup>13</sup> Due to its good performance, the MLS

method is often used in combination with other methods to construct shape function. In finite element analysis, the introduction of the MLS method enhances the shape function of the active element.<sup>14</sup> In the smoothed particle hydrodynamics (SPH), the MLS method was used to construct kernel functions to obtain higher consistency.<sup>15</sup> The MLS method obtains the local fitting coefficients through the weighted least squares (LS)<sup>16</sup> method in the support domain<sup>17</sup> based on the Gauss Markov error model, in which only the dependent variable contains errors.<sup>18</sup> The total least squares (TLS)<sup>19</sup> method is an estimation method based on the errors-in-variables (EIV) model. Unlike the Gauss Markov model, the errors of both independent variable and dependent variable are considered in the EIV model.<sup>20</sup> By replacing the LS estimation with the TLS estimation in the support domain, the MLS method is transformed into the moving total least squares (MTLS).<sup>21</sup>

Nevertheless, due to the impact of factors, such as the disturbances in the measurement environment, the measurement data often contains outliers that seriously deviate from the actual value.<sup>22,23</sup> The LS and TLS respectively used in the support domains of the MLS and MTLS method are not robust estimation methods.<sup>24,25</sup> When there are outliers in the domain, larger deviations will exist in the fitted values around outliers. Many studies have been carried out to reduce the negative impact from the outliers, with the proposed robust algorithms dividing into two main forms. One is to select a subsample from the discrete points to obtain the regression coefficient.<sup>26,27</sup> If the outlier is in the subsample, it will be automatically eliminated. However, this method may eliminate some of non-outliers, in which case the accuracy of approximation will be significantly affected.<sup>28</sup> The other type of algorithm is to identify outliers first and then weaken the influence of outliers by assigning weights to them. In this method, it is difficult to determine appropriate weights when multiple outliers with different levels exist in the discrete data.

In order to weaken the impact of outliers on reconstruction, we propose an improved MTLS (IMTLS) reconstruction method in this article. In the support domain, we deal with outliers by introducing an improved random sample consensus (RANSAC) and a correction parameter, and then the local approximation coefficients

are obtained based on the TLS estimation. RANSAC is a robust model estimation algorithm, especially when measurement data contains high proportion of outliers. However, it has limitations in data fitting influencing its accuracy and stability.<sup>29</sup> Therefore, RANSAC algorithm needs to be improved to estimate a relatively reliable initial model. The correction parameter associated with the random error and the node residual is introduced to detect and eliminate abnormal node.

The rest of this article is structured as follows: the second section is a brief introduction to the MLS, MTLS and RANSAC algorithm, the third section explains the principle and procedure of the proposed algorithm in detail, and the fourth part verifies the performance of the IMTLS method through numerical simulation and experimental data.

## 2. Introduction to the basic algorithms

### 2.1 The MLS method

Consider that there are  $n$  discrete points  $\mathbf{x}=[x_1, x_2, \dots, x_n]$  and corresponding  $\mathbf{y}=[y_1, y_2, \dots, y_n]$  in a bounded area  $\Omega$ . In the MLS method, the trial function  $f(x)$  for the fitting point  $x$  can be written as

$$f(x) = \sum_{j=1}^m p_j(x) a_j(x) = \mathbf{p}^T(x) \mathbf{a}(x) \quad (1)$$

where  $\mathbf{p}(x) = [p_1(x), p_2(x), \dots, p_m(x)]^T$  is a vector of the monomial basis function and the number of the basic is  $m$ , and  $\mathbf{a}(x)=[a_1(x), a_2(x), \dots, a_m(x)]^T$  is the coefficient vector to be solved. In this article, for 2D curve and 3D surface fitting, we choose  $\mathbf{p}(x)=[1, x]^T$  and  $\mathbf{p}(x)=[1, x, y]^T$ , respectively.

In order to solve the optimal  $\mathbf{a}(x)$ , the construct function  $E(x)$  is a quadratic function about  $\mathbf{a}(x)$ , i.e.

$$E(x) = \sum_{j=1}^n \theta(\|x - x_j\|/r) \left( \sum_{j=1}^m p_j(x_j) a_j(x) - y_j \right)^2 \quad (2)$$

where  $r$  is used to control size of the support domain, and  $\theta(\|x - x_j\|/r)$  is a non-negative and compactly supported weight function to attribute a weight to each node according to its position relative to  $x$ . The fitting property of MLS algorithm will be influenced by the weight function. For example, the fitting form will be interpolated when  $\theta(0)=\infty$ .<sup>30</sup> There are many kinds of weight functions with compact support such as exponential, Gaussian and cubic spline weight functions.

This article chooses exponential weight function, i.e.

$$\theta(s) = \begin{cases} \frac{\exp(-\alpha s^2) - \exp(-\alpha)}{1 - \exp(-\alpha)} & \text{if } s \leq 1 \\ 0 & \text{else} \end{cases} \quad (3)$$

where  $\alpha$  is a coefficient related to the convergence speed.

According to the principle of least squares, the coefficient vector  $\mathbf{a}(x)$  can be solved by

$$\frac{\partial E(x)}{\partial a_j(x)} = 2 \sum_{i=1}^N \theta(s) \left[ \sum_{j=1}^m p_j(x_i) a_j(x) - y_i \right] p_j(x_i) = 0 \quad (4)$$

where  $j=1, 2, \dots, m$ , solving equations (4) to obtain

$$\mathbf{a}(x) = \mathbf{A}^{-1}(x) \mathbf{C}(x) \mathbf{u} \quad (5)$$

where

$$\mathbf{u} = [y_1, y_2, \dots, y_N]^T$$

$$\mathbf{A}(x) = \sum_{i=1}^n \theta(\|x - x_i\|/r) \mathbf{p}(x_i) \mathbf{p}^T(x_i)$$

$$\mathbf{C}(x) = [\theta(\|x - x_1\|/r) \mathbf{p}(x_1), \theta(\|x - x_2\|/r) \mathbf{p}(x_2), \dots, \theta(\|x - x_n\|/r) \mathbf{p}(x_n)]$$

Taking equation (5) into equation (1), the function  $f(x)$  can be presented as

$$f(x) = \mathbf{p}^T(x) \mathbf{A}^{-1}(x) \mathbf{C}(x) \mathbf{u} \quad (6)$$

## 2.2 The MTLs method

Unlike the MLS using LS estimation to obtain the local fitting coefficients, the MTLs method applies the TLS estimation for those coefficients, in which the errors of both dependent variable and independent variables can be considered. In TLS, the solution of linear equation is considered, i.e.

$$(\mathbf{B} + \mathbf{E}_B) \mathbf{X} = \mathbf{L} + \mathbf{e} \quad (7)$$

where  $\mathbf{B}$  and  $\mathbf{L}$  are coefficient matrix and observation vector respectively,  $\mathbf{E}_B$  and  $\mathbf{e}$  are random error matrix of  $\mathbf{B}$  and  $\mathbf{L}$  respectively, and  $\mathbf{X}$  is the coefficient vector to be solved. According to the principle of total least squares, the solution of equation (7) can be transformed into a constrained optimization problem, i.e.

$$\mathbf{X} = \arg \min (\mathbf{e}_B^T \mathbf{e}_B + \mathbf{e}^T \mathbf{e}) \quad (8)$$

where  $\mathbf{e}_B = \text{vec}(\mathbf{E}_B)$ .

Therefore,  $\mathbf{X}$  can be solved by Lagrange multiplier method. The function is constructed as

$$\mathbf{J}(\mathbf{e}, \mathbf{e}_B, \boldsymbol{\varphi}, \mathbf{X}) = \mathbf{e}^T \mathbf{e} + \mathbf{e}_B^T \mathbf{e}_B + 2\boldsymbol{\varphi}^T (\mathbf{L} + \mathbf{e} - \mathbf{B}\mathbf{X} - \mathbf{E}_B \mathbf{X}) \quad (9)$$

where  $\boldsymbol{\varphi}$  is the vector of Lagrange multiplier. Let the first partial derivative of  $\mathbf{J}$  with respect to  $\mathbf{e}$ ,  $\mathbf{e}_B$ ,  $\boldsymbol{\varphi}$  and  $\mathbf{X}$  equal to 0, i.e.

$$\begin{cases} \frac{1}{2} \frac{\partial \mathbf{J}}{\partial \mathbf{e}} = \mathbf{e} + \boldsymbol{\varphi} = \mathbf{0} \\ \frac{1}{2} \frac{\partial \mathbf{J}}{\partial \mathbf{e}_B} = \mathbf{E}_B - \boldsymbol{\varphi} \mathbf{X}^T = \mathbf{0} \\ \frac{1}{2} \frac{\partial \mathbf{J}}{\partial \boldsymbol{\varphi}} = \mathbf{L} - \mathbf{B}\mathbf{X} + \mathbf{e} + \boldsymbol{\varphi} - \mathbf{E}_B \mathbf{X} = \mathbf{0} \\ \frac{1}{2} \frac{\partial \mathbf{J}}{\partial \mathbf{X}} = \mathbf{B}^T \boldsymbol{\varphi} + \mathbf{E}_B^T \boldsymbol{\varphi} = \mathbf{0} \end{cases} \quad (10)$$

derived from equations (10), we can obtain

$$\begin{cases} \mathbf{B}^T \mathbf{A} \mathbf{X} - \mathbf{B}^T \mathbf{L} = \mathbf{X} \boldsymbol{\mu} \\ \mathbf{L}^T \mathbf{L} - \mathbf{L}^T \mathbf{B} \mathbf{X} = \boldsymbol{\mu} \end{cases} \quad (11)$$

where  $\boldsymbol{\mu} = \mathbf{e}_B^T \mathbf{e}_B + \mathbf{e}^T \mathbf{e}$ .

The equations (11) can be transformed into

$$\begin{bmatrix} \mathbf{B} & \mathbf{L} \end{bmatrix}^T \begin{bmatrix} \mathbf{B} & \mathbf{L} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ -1 \end{bmatrix} = \begin{bmatrix} \mathbf{B}^T \mathbf{B} & \mathbf{B}^T \mathbf{L} \\ \mathbf{L}^T \mathbf{B} & \mathbf{L}^T \mathbf{L} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ -1 \end{bmatrix} = \boldsymbol{\mu} \begin{bmatrix} \mathbf{X} \\ -1 \end{bmatrix} \quad (12)$$

It can be seen from equation (12) that the problem of solving total least squares is transformed into obtaining the eigenvalues and eigenvectors of matrix  $\begin{bmatrix} \mathbf{B} & \mathbf{L} \end{bmatrix}^T \begin{bmatrix} \mathbf{B} & \mathbf{L} \end{bmatrix}$ . When  $\boldsymbol{\mu}$  is taken as the smallest eigenvalue of the matrix  $\begin{bmatrix} \mathbf{B} & \mathbf{L} \end{bmatrix}^T \begin{bmatrix} \mathbf{B} & \mathbf{L} \end{bmatrix}$ ,  $\mathbf{X}$  in the corresponding eigenvector is the solution of TLS. For the MTLs method, the construction of the fitting function is consistent with the MLS method, and the same weight function is used for the MTLs. The acquisition of local regression coefficients can be expressed as follows

$$\mathbf{D} = \begin{bmatrix} \mathbf{B}_w & \mathbf{L}_w \end{bmatrix}^T \begin{bmatrix} \mathbf{B}_w & \mathbf{L}_w \end{bmatrix} = \mathbf{M}^{-1} \boldsymbol{\Lambda} \mathbf{M} \quad (13)$$

where

$$\mathbf{L}_w = [\theta(\|x - x_1\|/r) y_1, \theta(\|x - x_2\|/r) y_2, \dots, \theta(\|x - x_n\|/r) y_n]^T$$

$$\mathbf{B}_w = [\theta(\|x - x_1\|/r) \mathbf{p}(x_1), (\|x - x_2\|/r) \mathbf{p}(x_2), \dots, \theta(\|x - x_n\|/r) \mathbf{p}(x_n)]^T$$

$$\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{m+1}), (\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{m+1})$$

$$\mathbf{M} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m+1}], \mathbf{D} \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

$$\mathbf{v}_i = [v_{i,1}, v_{i,2}, \dots, v_{i,m+1}]^T (i = 1, 2, \dots, m+1)$$

If  $\lambda_{m+1} \neq \lambda_m$ , combined with equation (12) and (13), the regression coefficient vector  $\mathbf{a}(x)$  is:

$$\mathbf{a}(x) = -\frac{1}{v_{m+1,m+1}} \left[ v_{m+1,1}, v_{m+1,2}, \dots, v_{m+1,m} \right]^T \quad (14)$$

In the MLS method, the coefficient matrix  $\mathbf{B}$  is error-free, while only the observation vector  $\mathbf{L}$  is contaminated. Nevertheless, this assumption is usually unrealistic in some fields of engineering, because the errors, such as sampling errors and instrument errors, are always brought into matrix  $\mathbf{B}$ . Therefore, the MTLs method can give a more reasonable approximation than the MLS method when both matrix  $\mathbf{B}$  and the vector  $\mathbf{L}$  contain errors.<sup>31</sup>

### 2.3 The RANSAC algorithm

The RANSAC algorithm, presented by Fischler and Bolles in 1981,<sup>32</sup> can robustly estimate the model parameters. It has been applied to many fields, especially in computer vision,<sup>33,34</sup> due to good performance in handling the data with a tremendous level of outliers and remarkably simple structure. The main idea of the RANSAC is to calculate parameters of hypothesized model by randomly sampling the subsample from the entire dataset,<sup>35</sup> and then the model is performed on the entire dataset. During this process, a distance threshold that is considered empirically<sup>36</sup> is set to calculate the inlier rate of each model, and sampling is stopped when the number of iterations reaches the defined value. The parameters corresponding to the highest inlier rate (the highest consensus) are selected as the model parameters of the total sample. The brief calculation steps of the RANSAC algorithm are described as follows:

---

#### Algorithm 1 The RANSAC method

---

**Input:** discrete data set  $C = \{x_i, y_i\}_{i=1}^N$

**Output:** model  $G(x)$

Set  $Q=0$

**for** ( $j=1$  to  $k$ ) **do**

    Randomly select a minimum subset from  $C$  to calculate model  $g(x)$  and get the residuals

    Count the number of inliers  $q$  of the current model according to distance threshold

**if** ( $q > Q$ ) **then**

$G(x) = g(x)$  and  $Q = q$

**end if**

**end for**

---

In order to get a subsample that is all inliers with probability  $p$ , the sampling times  $k$  can be obtained by the

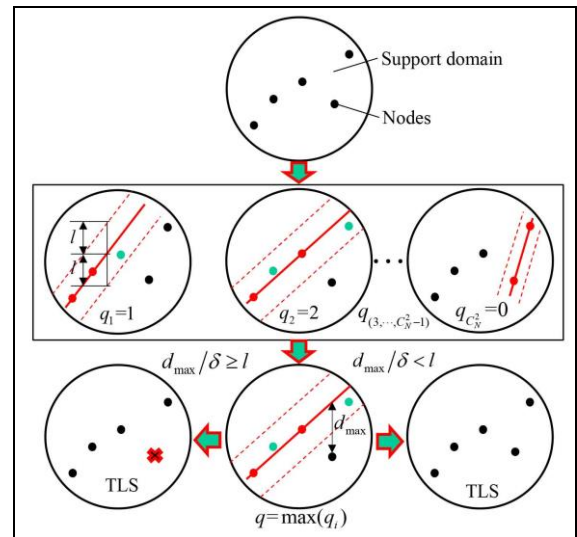
following formula,<sup>37</sup> i.e.

$$k = \frac{\log(1-p)}{\log(1-(1-p)^\zeta)} \quad (15)$$

where  $w$  is the proportion of outliers in the entire dataset, and  $\zeta$  is number of data nodes needed to calculate model parameters. For example,  $\zeta=2$  is taken in the linear parameter estimation. Then a reliable model is obtained through multiple iterations, and the node with the residual outside the distance threshold is considered as outlier.

### 3. The improved MTLs method

When there are outliers in the measurement data, the robustness and accuracy of both the MLS and MTLs method are not satisfactory due to their construction principles.<sup>38</sup> Therefore, this article proposes an IMTLs method to reduce the negative impact of outliers on reconstruction. In this method, an improved RANSAC and a correction parameter are introduced into the support domain of the MTLs method to detect and eliminate abnormal nodes, and then the local fitting coefficients based on the TLS estimation can be obtained. The improved RANSAC algorithm is different from the standard RANSAC algorithm in two aspects. Firstly, it considers every possible subsample to generate hypothesized models, which can enhance the stability of data reconstruction. Secondly, the threshold is automatically set to a value related to the random error of the data, which reduces the negative impact of the original method to set a threshold empirically.



**Figure 1.** The fitting procedure in the support domain of IMTLs method.

Then, we mainly introduce the principle and procedure of the algorithm reconstruction in support domain, as shown in Figure 1. Assuming that there are  $N$  nodes in single support domain, the reconstruction process of the MTLs method can mainly be divided into three steps. Firstly, an estimation model is obtained by the improved RANSAC algorithm. Secondly, the abnormal node is eliminated by introducing a correction parameter  $\delta$ . In this step, whether to eliminate the node with the largest residual is determined by the size of  $d_{\max}/\delta$  and  $l$ , where  $d_{\max}$  is the maximum residual and  $l$  is automatically set according to the random error of the data. Lastly, local approximation coefficients are determined by TLS estimation. In the support domain, the calculation process of the IMTLs method is shown as follows:

---

**Algorithm 2** The IMTLs method

---

**Input:** discrete data  $C=\{x_i, y_i\}_{i=1}^N$ , parameter  $\delta$  and  $l$

**Output:** local regression parameter  $\mathbf{a}$

Set  $Q=0$

Select different combinations including  $\zeta$  nodes from  $C$ , denoted as  $C_1, C_2, \dots, C_{C_v}^{\zeta}$

**for** ( $j=1$  to  $C_N^{\zeta}$ ) **do**

    Get model  $g(x)$  based on  $C_j$  and residuals  $\mathbf{d}=\{d_i\}_{i=1}^N$

    Count the number of inliers  $p$  of the model according to  $l$

**if** ( $q>Q$ ) **then**

$Q=q$  and  $\mathbf{D}=\mathbf{d}$

**end if**

**end for**

Get maximum residual  $d_{\max}=\max(\mathbf{D})$

**if** ( $d_{\max}/\delta \geq l$ ) **then**

    Eliminate the node with  $d_{\max}$

**else**

        Reserve the node with  $d_{\max}$

**end if**

Remaining data is processed based on TLS to get  $\mathbf{a}$

---

After eliminating the abnormal nodes, assume that  $t$  nodes  $\bar{\mathbf{x}} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_t]$  and  $\bar{\mathbf{y}} = [\bar{y}_1, \bar{y}_2, \dots, \bar{y}_t]$  are retained in the support domain. The solution of the TLS can also be obtained by the singular value decomposition (SVD).<sup>39</sup> As shown in the linear equation (7), perform the SVD on the augmented matrix

$$\mathbf{F} = [\mathbf{B} \ \mathbf{L}] = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \\ m+1 & t-m-1 \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma} \\ \mathbf{0} \end{bmatrix} \mathbf{V}^T = \mathbf{U}_1 \mathbf{\Gamma} \mathbf{V}^T \quad (16)$$

in which  $\mathbf{U}=[\mathbf{U}_1 \ \mathbf{U}_2]_{(t \times t)}$  and  $\mathbf{V}_{(m+1) \times (m+1)}$  is an orthogonal

matrix composed of eigenvectors of matrix  $\mathbf{F}\mathbf{F}^T$  and matrix  $\mathbf{F}^T\mathbf{F}$  respectively, and  $\mathbf{\Gamma}=\text{diag}[\zeta_1, \zeta_2, \dots, \zeta_{m+1}]$ ,  $\zeta_1 \geq \zeta_2 \geq \dots \geq \zeta_{m+1}$  are the singular values of matrix  $\mathbf{F}$ . Matrix  $\mathbf{\Gamma}$  and matrix  $\mathbf{V}$  can be presented as

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Gamma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_2 \end{bmatrix} \begin{matrix} m \\ 1 \end{matrix} \quad \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \begin{matrix} m \\ 1 \end{matrix}$$

When  $\mathbf{V}_{22}$  is non-singular and  $\zeta_m \neq \zeta_{m+1}$ , the solution of TLS is unique, i.e.

$$\mathbf{X} = -\mathbf{V}_{12} \mathbf{V}_{22}^{-1} \quad (17)$$

For the IMTLs method, the matrix  $\mathbf{F}$  is weighted, i.e.

$$\mathbf{F}_w = [\mathbf{B}_w \ \mathbf{L}_w] = \mathbf{U}_w \mathbf{\Sigma}_w \mathbf{V}_w^T \quad (18)$$

where

$$\mathbf{L}_w = [\theta(\|x - \bar{x}_1\|/r)\bar{y}_1, \theta(\|x - \bar{x}_2\|/r)\bar{y}_2, \dots, \theta(\|x - \bar{x}_t\|/r)\bar{y}_t]^T$$

$$\mathbf{B}_w = [\theta(\|x - \bar{x}_1\|/r)\mathbf{p}(\bar{x}_1), \theta(\|x - \bar{x}_2\|/r)\mathbf{p}(\bar{x}_2), \dots, \theta(\|x - \bar{x}_t\|/r)\mathbf{p}(\bar{x}_t)]^T$$

The solution expression of equation (17) is rewritten as

$$\mathbf{X}_w = -\mathbf{V}_{w12} \mathbf{V}_{w22}^{-1} \quad (19)$$

In the support domain of the IMTLs method, abnormal nodes are automatically detected and eliminated by the improved RANSAC method and the introduced correction parameter  $\delta$ . The outliers are evidently included in the abnormal node and can be effectively processed. The local fitting coefficients are obtained by the TLS estimation. As the movement of fitting point  $x$ , outliers and random errors are processed in the entire domain.

## 4. Case verification

In this section, both numerical and experimental cases were considered by the MLS, MTLs and IMTLs method, to verify the performance of the proposed method.

### 4.1 Case 1

Take the curve function

$$y = \sin(0.3\pi x) + 0.8\sin(0.5\pi x) \quad (20)$$

as a model to generate uniformly distributed discrete points  $(x_{0i}, y_{0i})$ . Outliers  $(0, \Delta y_i)$  and random errors of normal distribution with zero mean are added to  $(x_{0i}, y_{0i})$  to get simulation data  $(x_{1i}, y_{1i})$ . The reconstruction points

$(x_{2i}, y_{2i})$  are obtained after reconstruction. The sum of absolute differences can be presented as

$$s = \sum_{i=1}^n |y_{0i} - y_{2i}| \quad (21)$$

which is considered as the index to evaluate three algorithms.

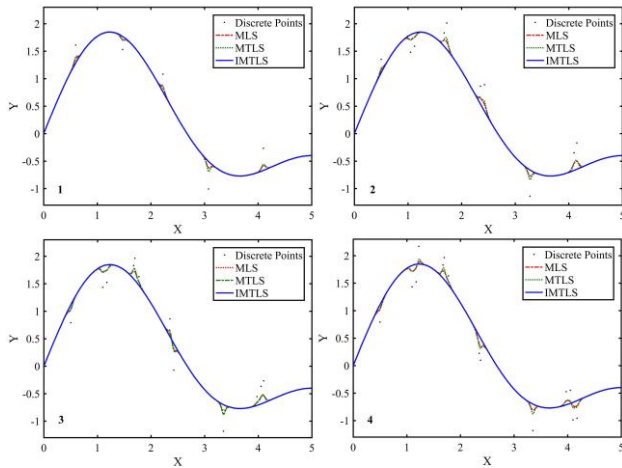
In this case, the parameter

$$l = \sqrt{\sigma_x^2 + \sigma_y^2} \quad (22)$$

Let  $n=201$  and  $r=[\max(x_0)-\min(x_0)] \times 4/100$ , where  $\min(x_0)=0$  and  $\max(x_0)=5$ . Under the same random error condition ( $\sigma_x=0.001$  and  $\sigma_y=0.001$ ), we add four different outliers to  $(x_{0i}, y_{0i})$ , and the  $s$  values obtained by three algorithms are shown in Table 1. The fitting curves are shown in Figure 2.

**Table 1.** The comparison of three methods in Case 1.

error parameter	MLS	MTLS	IMTLS
$s_1$	1.974703	1.995617	0.177981
$s_2$	3.752185	3.852139	0.177980
$s_3$	3.854453	3.878887	0.179741
$s_4$	3.906223	4.180141	0.180511
$\bar{s}$	3.371891	3.476696	0.179053
$\sigma_s$	0.808569	0.864746	0.001107



**Figure 2.** The curves fitting by three algorithms under different outlier conditions.

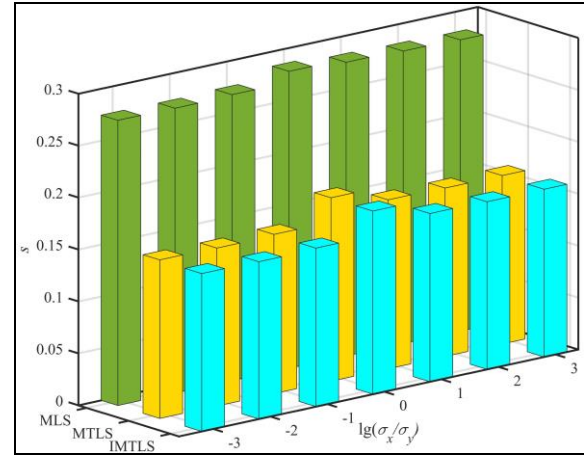
The figures and tables of the above cases show that outliers can be effectively detected and eliminated by the IMTLS method to reduce the incalculable negative impact on the reconstruction. Compared with the MLS and MTLS method, the error values of the reconstructed data processed by the IMTLS method is much smaller.

To further verify the performance of the IMTLS method, we take this case as an example, in which only

random errors are added in the simulation data. In Table 2, we can see that the error values obtained by the IMTLS fitting is still minimum among the three reconstruction methods, which shows that the IMTLS method also has good approximation performance under this situation, and Figure 3 shows the bar graph of error values under different random errors.

**Table 2.** The comparison of processing random errors in Case 1.

$\sigma_x$	$\sigma_y$	$s$		
		MLS	MTLS	IMTLS
0.000001	0.001	0.274568	0.152626	0.151917
0.00001	0.001	0.274228	0.152063	0.151351
0.0001	0.001	0.275371	0.153263	0.152555
0.001	0.001	0.285897	0.176635	0.176257
0.001	0.0001	0.282678	0.162684	0.161941
0.001	0.00001	0.281492	0.162066	0.161338
0.001	0.000001	0.280397	0.162179	0.161444



**Figure 3.** The  $s$  values under different random errors.

Furthermore, the CPU time of each algorithm is calculated by MATLAB based on the datasets with different size of  $n=201+i \times 200$  ( $i=0, 1, \dots, 4$ ). Two conditions are considered and presented in Table 3 and Table 4 ( $r_0=[\max(x_0)-\min(x_0)] \times 4/100$ , where  $\max(x_0)=5$  and  $\min(x_0)=0$ ) respectively. One is to fix the number of nodes in support domain, and the other one is to fix the size of support domain. In the first condition, CPU time of IMTLS method changes steadily with the increase of  $n$ . While in the second condition, the CPU time of IMTLS method increases rapidly as  $n$  increases because more nodes are included in single support domain. These results illustrate that the number of nodes in the support domain has a significant impact on the CPU time of IMTLS method. Therefore, in order to ensure the efficiency of IMTLS method in data reconstructing, the

size of support domain should be appropriately selected to control the number of nodes in support domain.

**Table 3.** The CPU times of three methods with a fixed number of nodes in support domain. (s)

$n$	201	401	601	801	1001
$r$	$r_0$	$(1/2)r_0$	$(1/3)r_0$	$(1/4)r_0$	$(1/5)r_0$
MLS	0.0148	0.0375	0.0477	0.0602	0.0813
MTLS	0.0075	0.0164	0.0250	0.0422	0.0477
IMTLS	0.1473	0.3148	0.4531	0.5977	0.7516

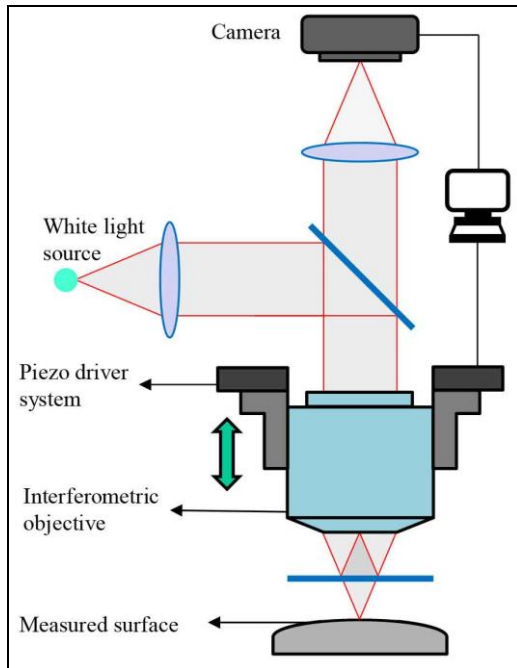
**Table 4.** The CPU times of three methods with a fixed radius of support domain. (s)

$n$	201	401	601	801	1001
$r$	$r_0$	$r_0$	$r_0$	$r_0$	$r_0$
MLS	0.0148	0.0563	0.1227	0.2407	0.4063
MTLS	0.0075	0.0273	0.0547	0.6000	1.1250
IMTLS	0.1473	0.9633	2.9602	45.3609	90.3750

## 4.2 Case 2

In this case, the surface data of standard ball was processed to verify the performance of the IMTLS method. The data is obtained by a commercial white light interferometer (WLI) - Taylor Hobson CCI 3000.

Figure 4 shows the schematic diagram of a WLI system. Firstly, a broadband illumination beam passes through an interferometric objective via a beam splitter. The beams that were reflected by reference mirror and measured surface were focused onto a camera. Interference fringes



**Figure 4.** Schematic diagram of a WLI system.

will generate when the optical path difference (OPD) between the reference and measurement arm is within the coherence length, and the visibility of the fringes increases as OPD decreases. A series of interferograms can be obtained through scanning the objective. Surface data can be obtained by tracking all coherence peaks or phase retrieval within the field of view of the objective.

The errors contained in experimental data will have a negative impact on obtaining the true profile of the ball measured, and the impact can be reduced by using appropriate algorithm for reconstruction. Three methods are used to reconstruct the experimental data respectively. Then, the reconstructed data are used for parameter regression based on simulated annealing algorithm<sup>40</sup> and the regressed radius is used as the evaluation index of reconstruction method. The radius of standard ball is calibrated as 14.402mm. Since the random error of measurement data cannot be calibrated technically,  $l$  is defined to the standard deviation of residuals fitted by the LS estimation in support domain. As shown in Figure 5, data selected in different locations are processed by three algorithms. As shown in Table 5, the radius corresponding to the IMTLS is more proximate to the calibrated value among the three algorithms.

**Table 5.** The regression radii of three methods in Case 2. (mm)

radius parameter	MLS	MTLS	IMTLS
$R_1$	14.340	14.435	14.405
$R_2$	14.441	14.393	14.408
$R_3$	14.723	14.632	14.613
$R_4$	14.615	14.617	14.420
$\bar{R}$	14.530	14.519	14.462
$\sigma_R$	0.148740	0.106425	0.087648

For the correction parameter  $\delta$  introduced by the IMTLS method, we set different values to observe the change of reconstruction result. As shown in Figure 6(a), when there are outliers in the data, the reconstruction result of the IMTLS method is obviously better than the other two algorithms. And the result will be consistent with that of the MTLS method as  $\delta$  increases, which shows that when  $\delta$  increases to a certain value, the node will not be eliminated in the support domain. As shown in Figure 6(b), when there are only random errors, the result of the IMTLS method is still better among the three algorithms. It is clear that  $\delta$  can be selected in this way for the curve and surface reconstruction.



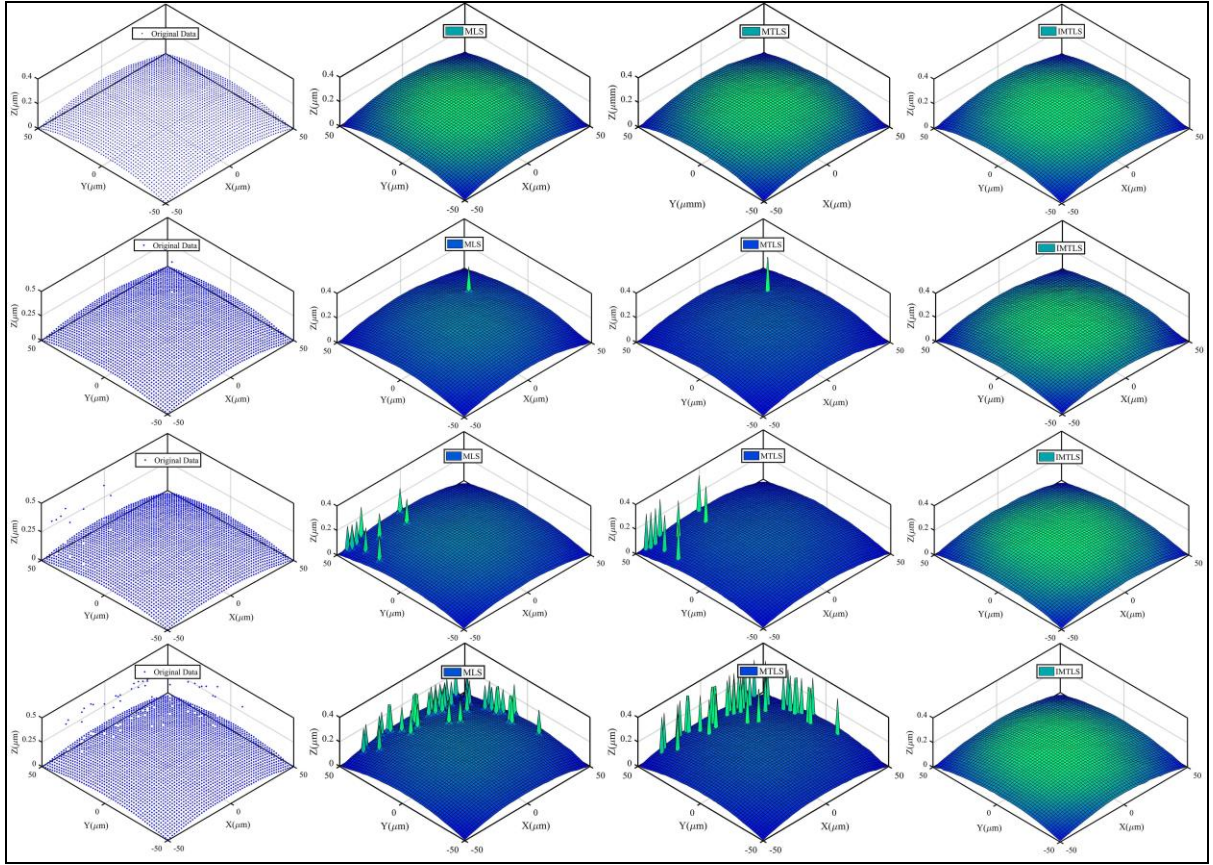


Figure 5. The processing of experimental data.

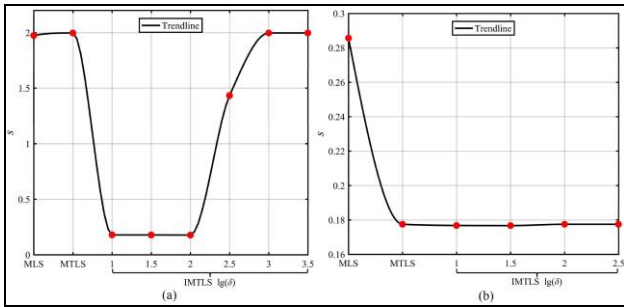


Figure 6. Trendlines of three algorithms in case 1: (a) the data of the fourth group, (b) the data only containing random errors.

Furthermore, we set different number of nodes that can be eliminated in the support domain, to observe the number of points eliminated in each support domain. The data of the fourth group in Case 2 is taken as an example.

As shown in the Figure 7, the  $l$  calculated by the residuals fitted by the LS estimation is relatively large in the support domain with outliers, while in the support domain without outliers, the obtained  $l$  is relatively small. Therefore, compared with the support domain without outliers, fewer points are eliminated in the support domain with outliers.

Figure 8 demonstrates the processing result with eliminated nodes, some outliers are not effectively

eliminated, because the number of eliminated nodes is not sufficient to eliminate all outliers in some support domains. In addition, the processed data are regressed by simulated annealing to get the regression radii, as shown in Table 6. When the outliers are effectively processed,

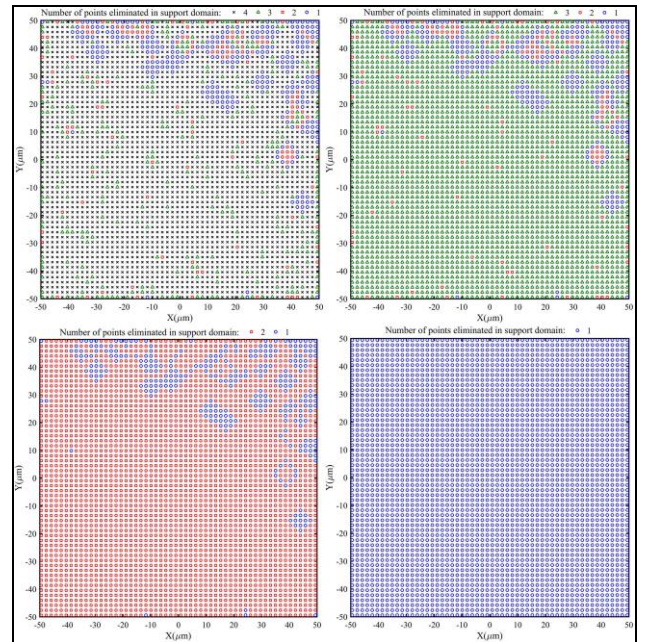


Figure 7. Number of points eliminated in each support domain under different conditions.



the corresponding regression radius is significantly closer to the calibrated value. This indicates that the outlier has a great negative influence on the parameter regression and can be processed effectively by setting an appropriate number of elimination nodes.

**Table 6.** The regression radii under different conditions. (mm)

$R_{44}$	$R_{43}$	$R_{42}$	$R_{41}$
14.420	14.394	14.464	14.484

**Figure 8.** The fitting surface under different conditions.

Through numerical simulation and experimental data verification, the IMTLS method shows better performance. As seen from Case 1, compared with the MLS and MTLs method, the IMTLS method can effectively deal with errors, regardless of whether there are outliers in the data. In addition, the IMTLS method also inherits the good local approximation properties from the MTLs method. We further verify the algorithm by processing measurement data of a standard ball, and the performance of proposed method is evaluated by regression radius. As seen from Case 2, the processed result by the IMTLS algorithm is more proximate to calibrated value.

## 5. Conclusion

The MLS and MTLs method show good performance in the fitting of discrete data, such as realizing effective approximation for local geometry feature and obtaining high-order continuous approximation functions with low-order basis functions, etc. However, these two reconstruction methods are not robust, because the outliers in the measurement data will have extremely negative impact on the fitting results. In order to reduce this impact, we proposed an improved MTLs method, in which an improved RANSAC and a correction parameter

are introduced into the support domain of the MTLs method to process the abnormal nodes, and then the local fitting coefficients are obtained by the TLS estimation. In this way, the improved MTLs method not only has the advantages of the MTLs method but also can effectively deal with outliers. Practically, we verified the proposed algorithm by dealing with experimental data obtained by CCI. The processing results of three cases show that the performance of the IMTLS method is significantly better than the other two algorithms.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 51605094) and the Fundamental Research Funds for the Central Universities (Grant No. WK2090050042).

## ORCID iD

Shuwen Lin, <https://orcid.org/0000-0002-1669-8840>.

## References

1. Mehrabi H and Voosoghi B. Recursive moving least squares. *Eng Anal Bound Elem* 2015; 58: 119–128.
2. Sun Y, Zhai J, Zhang Q, et al. An adaptive meshless method based on friction condition control strategy. *Eng Fail Anal* 2019; 106: 104148.
3. Wang Z, Qin X, Wei G, et al. Meshless method with ridge basis functions. *Appl Math Comput* 2010; 217(5): 1870-1886.
4. Arehpanahi, M, Vahedi, O. Modified weight function with automatic node generation in element-free Galerkin method for magnetic field computation. *IET Sci Meas Technol* 2015 9(8): 1043-1049.
5. Breitkopf P, Rassineux A, Savignat JM, et al. Integration constraint in diffuse element method. *Comput Method Appl M* 2004; 193(12): 1203-1220.
6. Mclain DH. Drawing contours from arbitrary data points. *Comput J* 1974; 4: 318-324.
7. Lancaster P and Salkauskas K. Surfaces generated by moving least squares methods. *Math Comput*

- 
- 1981; 37(155): 141-158.
8. Li Z, Feng, W, Yang J, et al. An investigation on modeling and compensation of synthetic geometric errors on large machine tools based on moving least squares method. *Proc IMechE, Part B: J Engineering Manufacture* 2018; 232(3): 412-427.
  9. Huang Z, Lei D, Huang D, et al. Boundary moving least square method for 2D elasticity problems. *Eng Anal Bound Elem* 2019; 106: 505-512.
  10. Harris MF, Kassab AJ and Divo E. A shock-capturing meshless scheme using RBF blended interpolation and moving least squares. *Eng Anal Bound Elem* 2019; 109: 81-93.
  11. Dabboura E, Sadat H and Prax C. A moving least squares meshless method for solving the generalized Kuramoto-Sivashinsky equation. *Alex Eng J* 2016; 55(3): 2783-2787.
  12. Fangyan F, Jiao L, Yanjin G, et al. Moving least squares particle hydrodynamics method for Burgers' equation. *Appl Math Comput* 2019; 356: 362-378.
  13. Kaplan AG and Kaplan YA. Developing of the new models in solar radiation estimation with curve fitting based on moving least-squares approximation. *Renew Energ* 2020; 146: 2462-2471.
  14. Mostafaiyan M, Wießner S and Heinrich G. Moving least-squares aided finite element method (MLS-FEM): A powerful means to predict pressure discontinuities of multi-phase flow fields and reduce spurious currents. *Comput Fluids* 2020; 211: 104669.
  15. Ghoneim AY. A smoothed particle hydrodynamics-phase field method with radial basis functions and moving least squares for meshfree simulation of dendritic solidification. *Appl Math Model* 2019; 77(2): 1704-1741.
  16. Yingjie Z and Liling G. A novel geometry error measurement methodology for coaxiality evaluation. *Proc IMechE, Part B: J Engineering Manufacture* 2021; 235(4): 627-639.
  17. Calvo R, Gómez E and Domingo R. Circle fitting from the polarity transformation regression. *Precis Eng* 2013; 37(4): 908-917.
  18. Wang B, Yu J, Liu C, et al. Data snooping algorithm for universal 3D similarity transformation based on generalized EIV model. *Measurement* 2018; 119: 56-62.
  19. Lei J, Liu S, Li ZH, et al. Image reconstruction algorithm based on the extended regularised total least squares method for electrical capacitance tomography. *IET Sci Meas Technol* 2008; 2(5): 326-336.
  20. Wang B, Liu C, Fang X, et al. A universally efficient algorithm and precision assessment for seamless 3D similarity transformation. *Meas Sci Technol* 2020; 31: 105006.
  21. Scitovski R, Ungar Š and Jukić D. Approximating surfaces by moving total least squares method. *Appl Math Comput* 1998; 93(2-3): 219-232.
  22. Wang C, D'Amato R and Gómez E. Confidence distance matrix for outlier identification: A new method to improve the characterizations of surfaces measured by confocal microscopy. *Measurement* 2019; 137: 484-500.
  23. Lou S, Tang D, Zeng W et al. Application of clustering filter for noise and outlier suppression in optical measurement of structured surfaces. *IEEE T Instrum Meas* 2020; 69(9): 6509-6517.
  24. Zheng S, Feng R and Huang A. A modified moving least-squares suitable for scattered data fitting with outliers. *J Comput Appl Math* 2020; 370: 112655.
  25. Ma Y, Liu S and Li Q. An advanced multiple outlier detection algorithm for 3D similarity datum transformation. *Measurement* 2020; 163: 107945.
  26. Čížek P. Least trimmed squares in nonlinear regression under dependence. *J Stat Plan Infer* 2006; 136(11): 3967-3988.
  27. Zhou G, Wang Q and Xiao Z. Robust outlier removal using penalized linear regression in multiview geometry. *Neurocomputing* 2017; 267: 455-465.
  28. Amir A and Levin D. Quasi-interpolation and outliers removal. *Numer Algorithms* 2018; 78: 805-825.
  29. Raguram R, Chum O, Pollefeys M, et al. USAC: A Universal Framework for Random Sample Consensus. *IEEE Trans Pattern Ana Mach Intel* 2013; 35(8): 2022-2038.
  30. Sober B, Aizenbud Y and Levin D. Approximation of functions over manifolds: A Moving Least-Squares approach. *J Comput Appl Math* 2021; 83: 113140.

- 
31. Wang H and Xiang H. Quantum algorithm for total least squares data fitting. *Phys Lett A* 2019; 383(19): 2235-2240.
  32. Fischler MA and Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 1981; 24: 381-395.
  33. Yu L, Bi Q, Ji Y, et al. Vision based in-process inspection for countersink in automated drilling and riveting. *Precis Eng* 2019; 58: 35-46.
  34. Jeon MH, Khambampati AK, Kim BS, et al. Image reconstruction in EIT with unreliable electrode data using random sample consensus method. *Meas Sci Technol* 2017; 28: 055403.
  35. Zheng Y, Liu J, Liu Z, et al. A primitive-based 3D reconstruction method for remanufacturing. *Int J Adv Manuf Technol* 2019; 103: 3667–3681.
  36. Nejad ZH and Nasri M. A-RANSAC: Adaptive random sample consensus method in multimodal retinal image registration. *Biomed Signal Proces* 2018; 45: 325-338.
  37. Civera J, Grasa OG, Davison AJ, et al. 1-Point RANSAC for extended Kalman filtering: Application to real-time structure from motion and visual odometry. *J Field Robot* 2010; 27: 609–631.
  38. Gu T, Tu Y, Tang D, et al. A robust Moving Total Least-Squares fitting method for measurement data. *IEEE T Instrum Meas* 2020; 69(10): 7566-7573.
  39. Zhang JQ and Ovaska SJ. ADC characterization based on singular value decomposition. *IEEE T Instrum Meas* 2002; 51(1): 138-143.
  40. Tamizharasan T, Barnabas JK, Pakkirisamy V. Optimization of turning parameters by using design of experiments and simulated annealing algorithm based on audible acoustic emission signals. *Proc IMechE, Part B: J Engineering Manufacture* 2012; 226(7): 1159-1173.