International Journal of the Physical Sciences Vol. 5(11), pp. 1641-1646, 18 September, 2010
Available online at http://www.academicjournals.org/IJPS
ISSN 1992 - 1950 ©2010 Academic Journals

Full Length Research Paper

Unilateral left-side quartile algorithm based data processing scheme for 3D scattered point data

Huipu Xu*, Ying Hu, Zi Ma and Jianhui Nie

Automation Research Center, Dalian Maritime University, Dalian 116026, China.

Accepted 20 August 2010

Currently, the data processing approaches for 3D point cloud data are based on the topology of the data in reverse engineering. A novel data processing scheme for the 3D measurement results with scattered, unorganized and unordered 3D point data is presented. An improved algorithm, namely, unilateral left-side quartile method is put forward in this paper based on the frequency method and quartile in statistics. According to the presented algorithm, the point data are firstly portioned in space, based on frequency algorithm. Then the threshold of the noise data can be found from large number of data. K-nearest points can be found through the K-Dimensional binary search tree (K-D tree) established based on 3D point cloud, and several nearest border upon points (NBUP) are defined around each noise point. By means of the NBUP, the noise point data can be recovered. Finally, the effectiveness in reverse engineering of the proposed scheme is demonstrated via a testing result based on Handyscan 3D scanner.

Key words: Unilateral left-side quartile method, frequency method, unorganized point cloud, data preprocessing, K-D tree.

INTRODUCTION

There are more mature filtering technologies for point cloud with regular arrangement or clear topological relationship (Xiaoping et al., 2001; Yangyu et al., 2006; Haber et al., 2001; Feng, 1998; Simon, 2004; Sun et al., 2001; Charles, 1992; Blanchard et al., 1998) for scan-line point cloud with all data in the same scan line in a plane, there are some certain rule to follow and some approaches at present (Xiaoming and Kangping, 2004). The filtering algorithm is presented in this paper for the scattered and disordered point cloud without obvious geometric distribution characteristics, which are obtained through random sampling. The point cloud coming from CMM under random scanning, laser-point measurement system, robot vision laser measurement system (Zi et al., 2006) through multi-piece and multi-angle scanning are mentioned above.

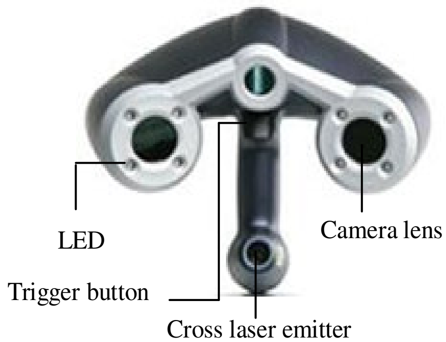
Lihui and Baozong (2008) and Schall et al. (2005) proposed filtering algorithms on the entirely discrete point cloud, in which the former, first of all, uses the Fuzzy C-means clustering to remove the large-scale noise, and then bilateral filtering method was used to remove the small-scale noise, and so this method requires two-step to remove the noise completely; the latter has some difficulties to be used in practice because it requires a more profound mathematical theory as support. Now there is no easy way that can be used for reference on data processing of 3D discrete point cloud.

A simple and practical unilateral left-side quartile algorithm on noise determination is puts forward in this paper. Based on this algorithm, the point data are smoothed and the validity and usefulness of the method are proved based on specific test platform.

HANDYSCAN 3D LASER SCANNER

Here, the application background of the proposed algorithm in this paper will be introduced briefly. The point data used in filtering the proposed algorithm comes from Handyscan 3D laser scanner (Figure 1) produced by Creaform company in Canada.

*Corresponding author. E-mail: hpx1212@163.com. Tel: (+86)0411-84725191.



(a)

(b)

Figure 1. Handayscan 3D laser scanner. (a) main view; (b) Top view.

Handyscan 3D laser scanner uses self-positioning technology, and do not need other assistive tracking or positioning devices, for example, additional tracking device or measuring arm of CMM determines its location or position relationship relative to the measured objects. During the scanning process, the scanner can capture the target points in real-time, and then compute and record its relative positive to the scanner. The scanning object can be positioned by these target points, and the 3D size information of the object can be acquired by the 3D laser after positioning finally, and the information can be displayed on host computer in real time. At the same time, the user can move the scanner to any position freely during the scanning process by using these target points, and the scanning object can also be moved. The aforementioned advantage makes it very easy for the scanner to scan any size, shape and internal and external scanning. The system components are shown in Figure 2.

ALGORITHM

Algorithm is divided into the following three steps:

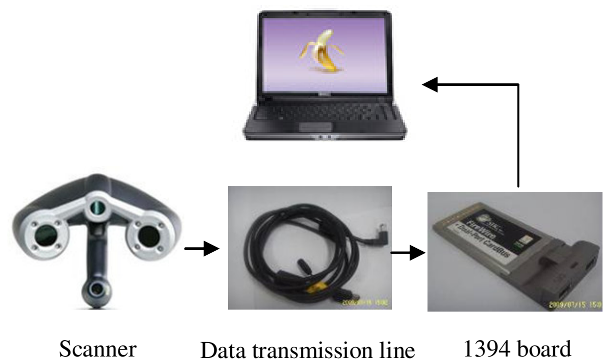


Figure 2. System components.

1. The region segmentation is completed on the origin point cloud using frequency method in probability and statistics, and the tab is set for each region.
2. The noises in each region are found according to unilateral left-side quartile algorithm.
3. The noise data are deleted and the real information in the position of noise data is restored.

The first two steps can be considered as a process of looking for noise data.

Determination of noise data

According to the point data, which includes noise data, the constraint space that embraces all of the data can be established. A reasonable way to determine constraint space is as follows:

1. The origin 3D point cloud should not only include the normal data signal, but also should be mixed with discrete noise and system noise.
2. The system noise is deleted directly from data file before determining constraint space, so, the inappropriate choice of constraint space can be avoided because of the exiting of system noise. The noise filtering algorithm is thus summarized.

1. Deleting the system noise: According to the size of measured object, 3D space Ω_1 with a certain redundancy is determined in advance, and the system noise falling out of Ω_1 are deleted.
2. Partitioning of the point cloud set into regions: A minimum bounding box which can include all the point data is chosen in the remaining data Ω_2, and the maximal and the minimal values of X, Y, Z coordinate, are defined as x_min, x_max, y_min, y_max, z_min和 z_max, respectively. The following 8 points are used as the vertexes of the bounding box Ω_2:

A(x_min,y_min,z_min) B(x_max,y_min,z_min)
C(x_max,y_min,z_max) D(x_min,y_min,z_max)
E(x_max,y_max,z_max) F(x_min,y_max,z_max)
G(x_min,y_max,z_min) H(x_max,y_max,z_min)

The six surfaces of the bounding box are ABCD, ECBH, ABGH, ADFG, EFGH and CDEF (Figure 3).

According to the coordinate range of discrete point cloud, the bounding box Ω_2 is divided into m, n, p interval in x-direction,

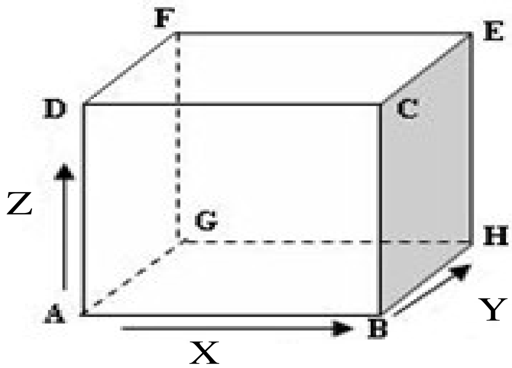


Figure 3. The sketch map of total restriction space.

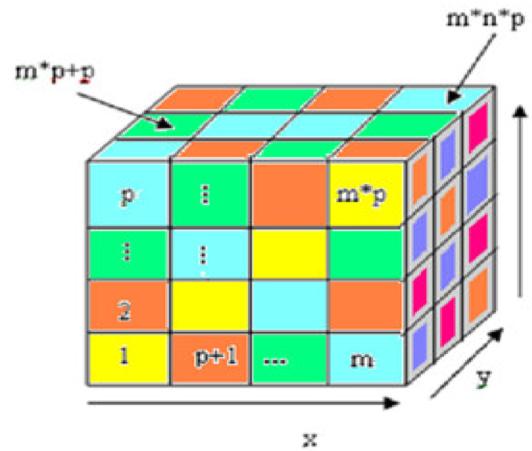


Figure 4. The sketch map of sub-restriction space.

y-direction and z-direction, respectively. The computation of interval number refers to the experience formula proposed by Sturges in America (1926):

$$k = 1 + 3.32 * \lg N \tag{1}$$

where k the is interval number of data partition and N is the number of points. To ensure that the interval number in x-direction, y-direction, z-direction are integer and the interval number of partition is more than that of experience, the following improved formulas are used to compute m, n, p:

$$\begin{aligned} m &= [1 + 3.32 * \lg L] + 1 \\ n &= [1 + 3.32 * \lg L] + 1 \\ p &= [1 + 3.32 * \lg L] + 1 \end{aligned} \tag{2}$$

where L is the number of sample data points, [ ] represent the rounding operation, therefore, the class interval in x-direction, y-direction and z-direction, respectively, are:

$$\begin{aligned} \text{int}\_m &\approx (x\_\max - x\_\min)/m \\ \text{int}\_n &\approx (y\_\max - y\_\min)/n \\ \text{int}\_p &\approx (z\_\max - z\_\min)/p \end{aligned} \tag{3}$$

Accordingly, there are $m*n*p$ sub-rounding boxes, and each sub-rounding box is tagged with $\omega_{i,j,k}$. The scope of each sub-rounding box is determined according to the above formulas, and the position of each point in the whole rounding box will be determined, where $i \in [1,p]$, $j \in [1,m]$, $k \in [1,n]$.

Each sub-rounding box is marked in accordance with that shown in Figure 4. The frequency of data points falling into each sub-rounding box is:

$$f_{i,j,k} = \frac{n_{i,j,k}}{N} \tag{4}$$

where $i \in [1,p]$, $j \in [1,m]$, $k \in [1,n]$, $N = m*n*p$, $n_{i,j,k}$ is the number of data points falling into sub-rounding box $\omega_{i,j,k}$.

3) Looking for noise data in partition region according to unilateral left-side quartile algorithm: Where $f_{i,j,k}$ is frequency of falling into sub-rounding box $\omega_{i,j,k}$. If $f_{i,j,k}$ is larger, the data points of falling into sub-rounding box with same size is more, and it can be shown that these points are real data points. On the contrary, If $f_{i,j,k}$ is smaller, the data points of falling into sub-rounding box with same size is less. According to the impossibility principle of small probability event in practice, it is thought as a small probability event when $f_{i,j,k}$ is smaller, and an impossible event in an experiment, when the data points falling into the sub-rounding box are considered as untrue points, namely noise data. Then how to determine the value of $f_{i,j,k}$ so that the situation can be considered as a small probability event is a crucial point here. Therefore, a unilateral left-side quartile algorithm on the threshold of $f_{i,j,k}$ is proposed in this paper.

As the complexity of object's surface and homogeneity of data points measured, $f_{i,j,k}$ is 0 for some sub-rounding box. The frequency of data points falling into each sub-rounding box $f_{i,j,k}$ can be thought as a random variable, and the value of $f_{i,j,k}$ which is equal to 0 needs to be deleted before using the proposed algorithm in this paper. The remaining variables are arranged in ascending order, however, some parameters such as the median M, upper quartile F_U, lower quartile F_L, dispersion of quartile dF, elimination point ρ and so on, are then computed. On the basis of traditional quartile principle, if variable falls into the following interval the variable is considered as invalid data:

$$\left| f_{i,j,k} - M \right| > \beta \cdot dF \tag{5}$$

where β is the constant, and its value can be adjusted depending on the measurement precision of system, which is desirable at 2, 1, and 0.5, etc.. The elimination point ρ is defined as:

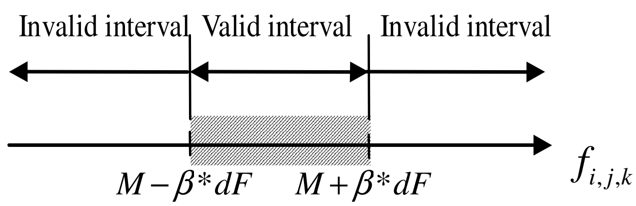


Figure 5. The elimination interval schematic of quartile method.

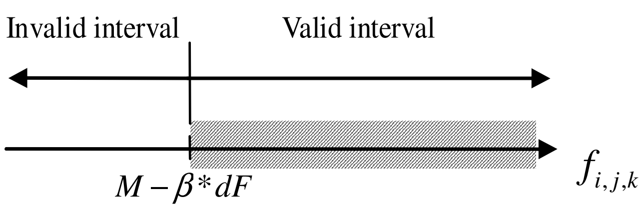


Figure 6. The elimination interval schematic of unilateral left-side quartile method.

$[M-\beta * dF, M+\beta * dF]$, and the data that are located outside the elimination point are taken as the invalid data. The valid interval is shown in Figure 5.

According to the background of data in this paper (the variables $f_{i,j,k} \in R^{+}$), the data, which are in the right side of the Figure 5 can not be regarded as invalid data. Therefore, the elimination point ρ is modified as $[M-\beta * dF, 1)$; accordingly, the invalid interval is modified as follows:

$$f_{i,j,k} - M < -\beta \cdot dF \tag{6}$$

Namely,

$$f_{i,j,k} < M - \beta \cdot dF \tag{7}$$

The modified invalid interval is shown in Figure 6.

If the frequencies of the data falling into sub-rounding box $f_{i,j,k}$ are being able to locate the invalid interval shown in Equation 1, these data are considered as noise data, and the depth value in these points need to be modified.

Restoring the information on the noise data

If the noise data are deleted directly from the data file, some important features of the measured object will be lost. Therefore, the blankness, as a result of deleting the noise data, must be filled. The restoring plan of data information on noise data is as follows:

It is an effective means of restoring the information of blankness, which occur as a result of deleting the noise data according to the neighbor information around the noise data. In looking for neighbor points, generally, the Euclidean distances of each point to other points are calculated and sorted in ascending order, and then the k smallest points are chosen. But for such huge point cloud, the efficiency of looking for neighbor points in this way is very low, and

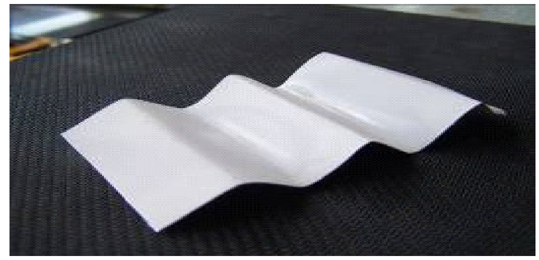


Figure 7. Object.

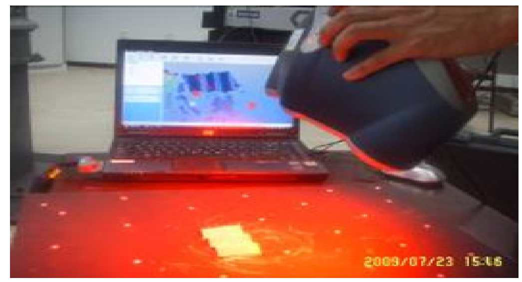


Figure 8. Object in scanning.

it is not necessary to calculate all the distances from one point to another. In order to speed up neighborhood search, the hierarchy fitting for neighborhood search is established: k-d tree.

Multi-dimension binary search tree (for short, K-D tree) is a spatial decomposition data structure object-based in organization (Friedman et al., 1977; Bentley, 1975; Gaede and Gunther, 1998; Samet, 1990). Partition value of one dimension is stored in the internal nodes of the tree, and all data set whose values are less than or equal to partition value correspond to its left sub-tree, and the other data set correspond to its right sub-tree. Leaf nodes are used to store certain size of data set. In general, the dimension is chosen in turn, and the median value of nodes is taken as partition value. The root node of the tree corresponds to the set of all points in k-dimension space. For *i*th layer of the tree, d-dimension is selected as identifier where j = (i mod k) +1. The set is partitioned into two sub-sets using the median node in d_i-dimension, which correspond to children nodes of root node. The above division process is repeated until each leaf node corresponds to a certain size of data space.

When K-D tree research algorithm is used, first, the data set in 3D space are expressed as data structure based on K-D tree, as such, an optimized K-D tree in research space is created; then, the k-nearest points in a certain distance are searched based on K-D tree.

When k-nearest points around blank point are searched, the median of k-nearest points is taken as the value in the position of blank point.

RESULTS

The effectiveness of the proposed algorithm in this paper is validated through the experiment, and the experiment is concluded based on Handyscan 3D laser scanner introduced here.

Figure 7 shows the wave object, the object in scanning is shown in Figure 8, while the scanning result is demonstrated in Figure 9, and the 30856 data points are

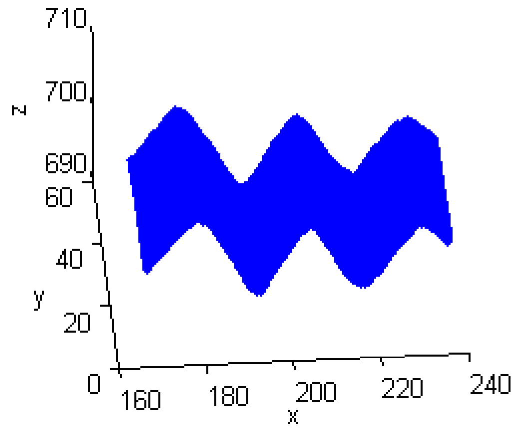


Figure 9. The origin data.

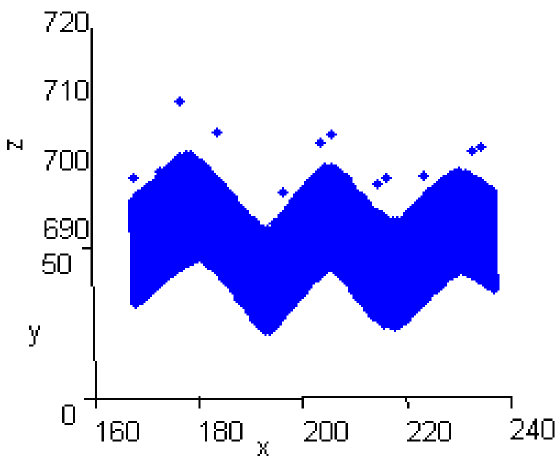


Figure 10. The data with noise.

included. The data points are smooth because the preprocess is finished in the process of scanning. In order to validate the effectiveness of the algorithm presented in this paper, the data in Figure 9 are processed and polluted with random noise. The data after pollution is as shown in Figure 10.

Above all, noise data are searched according to the algorithm introduced in here. By calculating, $m = n = p = 17$, β is taken as 1, and the threshold of $f_{i,j,k}$ is 0.0097, hence, 38 noise data are found; the noise data found are marked with red (Figure 11). After the noise data have been deleted, the information of blank points in the position of the noise data are restored by the algorithm introduced in here. Firstly, k-d tree is established based on 3D data points; secondly, k-nearest neighbor points of the noise data, in which k is taken, 8 are searched according to k-d tree; thirdly, the blank point is replaced with the median of 8-nearest neighbor points. The data points after noise preprocessing are shown in Figure 12.

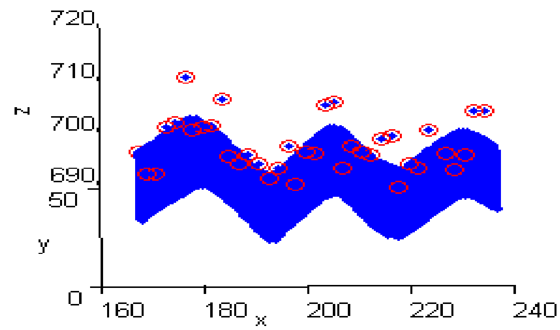


Figure 11. The found noise.

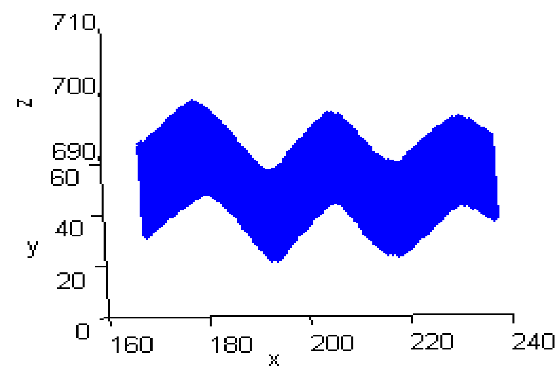


Figure 12. The data after restoring

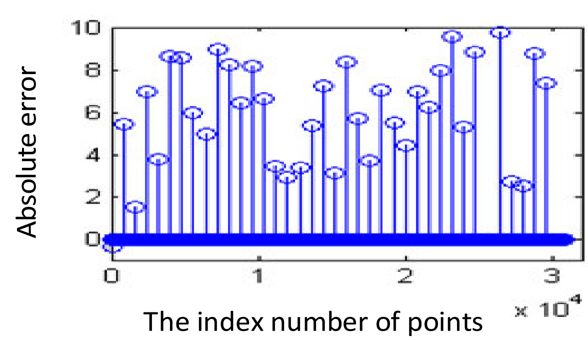


Figure 13. The absolute error.

Finally, the point data before (Figure 9) and after filtering (Figure 12) are compared, and the absolute error and relative error are shown in Figures 13 and 14, respectively, in which the maximal relative error is 1.38%.

Conclusion

A unilateral left-side quartile algorithm based on probability on the smoothing problem of 3D discrete point data is put forward in this paper. The algorithm is

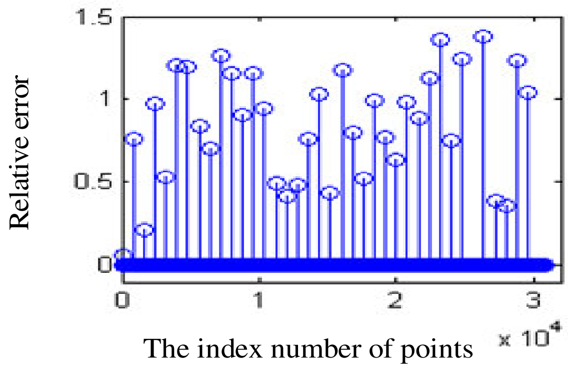


Figure 14. The relative error.

validated based on the test platform of Handyscan 3D laser scanner. From the results, the noise data are positioned accurately, and the effectiveness of smoothing is good. After smoothing, the point cloud not only maintains the details of object's features, but also the efficiency of smoothing is high. The limitations that the topology of 3D data must be known or the arrangement of 3D data must be regularly faced by other smoothing algorithm are broken, and a more efficient and practical data pre-processing strategies for 3D surface reconstruction is provided by this algorithm. Furthermore, this algorithm is simple and easy to operate in practical application.

REFERENCES

Bentley JL (1975). Multidimensional binary search trees used for associative searching [J]. Comm. ACM, 18(9): 509-517.

Blanchard AJ, Gader P, Correa AC, Hocaoglu AK (1998). The Use of Spline Based Wavelet Filtering to Improve Classification[C]//Processing of SAR Imagery, Seattle, 1757-1759.

Charles C (1992). An introduction to wavelets [M]. Boston: Academic Press, pp. 20-30.

Feng G (1998). Data smoothing by cubic spline filters [J]. IEEE Trans. on Signal Processing, 46: 2790-2796.

Friedman JH, Bentley JL, Finkel RA (1977). An algorithm for finding best matches in logarithmic expected time [J]. ACM Trans. on Mathematical Software, 3(3): 209-226.

Gaede V, Gunther O (1998). Multidimensional access methods [J]. ACM Computing Surveys, 30(2): 170-231.

Haber J, Zeilfelder F, Davydovt O, Seidel HP (2001). Hans-Peter Seidel, Smooth Approximation and Rendering of Large Scattered Data Sets[C]// Proceedings of IEEE Visualization 2001, San Diego, pp. 341-347.

Lihui W, Baozong Y (2008). Robust Fuzzy C-M eans and Bilateral Point Clouds Denoising [J]. J. Beijing Jiaotong University, 32(2): 18-21.

Samet H (1990). The design and analysis of spatial data structures [M]. Massachuset: Addison-Wesley, pp. 188-213.

Schall O, Belyaev A, Seidel HP (2005). Robust Filtering of Noisy Scattered Point Data. Emgraphics Symposium on Point-Based Graphics, pp. 71-77.

Simon D (2004). Data Smoothing and Interpolation Using Eighth-order Algebraic Splines [J]. IEEE transactions on signal processing, 52(4): 1136-1144.

Sturges H (1926). The choice of a class-interval [J]. Am. Stat. Assoc., 21: 65–66.

Sun W, Bradley C, Zhang YF, Loh HT (2001). Cloud data modeling employing a unified non-redundant triangular mesh [J]. Comput. Aided Des., 33(2): 183–193.

Xiaoming D, Kangping Z (2004). A Random Filter Algorithm for Reducing Noise Error of Point Cloud Data [J]. J. Image Graphics, 9(2): 245-248.

Xiaoping H, Xiaoming D, Youlun X (2001). Modeling Technique in Reverse Engineering. China Mech. Eng., 12(5): 539-542.

Yangyu P, Dongbo L, Fei T (2006). Noise reducing of data based on wavelet technology [J]. J. Mach. Design, 23(1): 31-41.

Zi M, Huipu X, Ying H, Jin H (2006). Robot Vision System and Artificial Neural Network for Model Reconstruction in Reverse Engineering [C]// Proceedings of the 6th World Congress on Intelligent Control and Automation, Dalian, pp. 9073-9078.