Research on Isolated Cloud Point Fast-Filtering Algorithm for Digital Twin

Hongjun Li¹,², Xiaorui Guo¹,², Kang Yang¹,², Chi Zhang¹ and Zhihan Zhang¹,²∗

¹ School of Mechanical Engineering and Automation, Wuhan Textile University, room N107, Nanhu campus, Wuhan, China
² Industrial Detonator Intelligent Assembly Hubei Engineering Research Center, room N107, Nanhu campus, Wuhan, China

Email: 1051093701@qq.com

Abstract. With the establishment of the digital twin stereoscopic warehouse concept, the importance of twin data has become increasingly prominent for digital twin system. Aiming at the problem of low accuracy in obtaining twin data from warehouse stocks, an isolated point cloud filtering algorithm combined with digital signal processing is proposed. The algorithm can retrieve the coordinate value of isolated point in original twin point cloud by constructing a twin point cloud fitting model, thereby filter out the point information of isolated region and obtain the warehousing twin data. The experiment results show that the algorithm can filter all isolated points while keeping the characteristics of original twin point cloud. The method provides accurate twin data support for digital twin stereoscopic warehouse.

1. Introduction

Digital twin uses information, such as physical models, data acquisition, historical data and so on, to achieve high-fidelity virtual mapping from physical space to digital virtual space[1]. It was proposed by Grieves[2] in 2003. With the information wave represented by big data, internet of things, virtual reality, cloud computing and other technologies sweeping the world[3], the concept of digital twins has been introduced into intelligent warehousing, which has profoundly affected the planning, construction and governance of warehousing. In 2019, Tao [4] proposed ten major applications of digital twin, the theoretical model of digital twin stereoscopic warehouse was built for the problems of low warehouse utilization and throughput to be improved. Obtaining complete and accurate twin data from physical entities is a significant cornerstone for constructing the five-dimensional digital twin warehouse stereoscopic model.

As a new equipment of active three-dimensional spatial information acquisition sensor, LiDAR determines the distance information between the laser source and the target by detecting light pulses or modulated pulses[5], which has the advantages of high efficiency and accuracy. And it has been widely applied in the surveying and mapping fields such as stockpiling measurement, hydrological modeling, and forest investigation [6]. Due to the complexity of spatial environment, noise is inevitably introduced into the twin point cloud. Isolated points, without any connection, is the typical noise in twin point cloud, which were caused by the occlusion of non-target objects.

In order to solve the isolated points in the original twin point cloud, we propose a twin point cloud filtering algorithms which is combined with digital signal processing technology. Under the premise of maintaining the main characteristics of twin point cloud, the algorithm filters out all isolated points and provides accurate twin data for the system of digital twin stereoscopic warehouses.
2. Principle of Isolated Data Filtering Algorithm

2.1. Algorithm Framework
The algorithm base on the feature that each twin points belong to same twin point cloud frame collected by LiDAR have the same X coordinate value in the Cartesian coordinate system. The adjacent twin points height difference series as a discrete non-periodic time domain signal, use discrete Fourier transform from time domain to frequency domain to obtain the twin point cloud fitting model. When we solve the twin point cloud fitting model, we use it to retrieve the position information of isolated points from original twin point cloud. Isolated points in the twin point cloud can be filtered out quickly and accurately. The algorithm is mainly divided into three steps: twin point cloud preprocessing, twin point cloud fitting model establishing and isolated point retrieving. The overall process as follows:

1) A twin point cloud frame(n twin points) are extracted from the original twin point cloud, and the height difference of adjacent twin points (difference of Z coordinate value) is calculated according to the order of Y coordinates. Compose a data set \( \Delta z_{(n-1)} \).

2) \( \Delta z_{(n-1)} \) is used discrete Fourier transform to deal with the height difference data set, and the characteristic frequency signal caused by isolated points are filtered out after converting the time domain signal to the frequency domain signal. The inverse Fourier transform is performed on the low-pass filtered frequency domain signal to obtain the filtered time domain signal, which is restored height information to construct a twin point cloud fitting model.

3) Retrieve the position coordinates of the isolated point in the original twin point cloud through the twin point cloud fitting model, and filter out the point information of the isolated area to obtain the twin data that does not contain the isolated points.

The details of the specific steps in the processing flow are shown in figure 1:

![Flow chart of isolated data filtering processing](image)

Figure 1. Flow chart of isolated data filtering processing.

Twin point cloud preprocessing (Step 1): According X coordinate value, algorithm extract a twin point cloud frame from the twin point cloud data file. The bubble sort method is used to reorder the Y coordinate value from small to large, and Z coordinate value is corresponded with original Y coordinate value. So we obtain the height data set \( z_{(n)} \). Set index number for height difference between two adjacent point clouds.

Twin point cloud fitting model establishing (step 2): Convert of height difference data set \( \Delta z_{(n-1)} \)
from time domain signal to frequency domain signal by discrete Fourier transform. A low-pass filter is constructed to filter out high frequency signals in frequency domain. The time domain signal \( \Delta z_{(n-1)} \) is restored by inverse Fourier transform. All Z coordinate value of the fitting model point are derived by previous basis point, and the first basis point is \( z_{(1)} \).

Isolated point retrieving (Step 3): Difference between the twin fitting model in step 2 and the original twin point cloud. If the difference value exceeds the threshold, the point is treated as an isolated point and filtered from the original twin data. Finally, we obtain the target twin data.

2.2. Twin Point Cloud Preprocessing
At the scanning process of vehicle LiDAR, Twin points in the twin point frame have the same X coordinate value. Therefore, the single-frame twin point cloud can be extracted by retrieving twin points with same X coordinate value from original twin point. Due to the existence of occlusions in the actual measurement process, the data storage order cannot be carried out according with the metabolic direction of the Y coordinate value. If the subsequent algorithmic steps is directly carried out, isolated data information in the fitting model cannot be filtered. Therefore, we can obtain a twin point cloud frame \( P_{(n)} = (x, y, z_{(n)}) \) consist of n twin points which had been sorted according to the Y coordinate value. Based on \( z_{(1)} \), (figure 2(a)), the height difference series is obtained as follows:

\[
\Delta z_{(m)} = z_{(m+1)} - z_{(m)}, 1 \leq m \leq n - 1
\]  
\[
(1)
\]

2.3. Twin Point Cloud Fitting Model Establishing
The purpose of this processing step is obtaining twin point cloud fitting model without isolated points. Because the isolated points is far away from the subject of twin point cloud in the spatial domain, the data fluctuates violently in the height difference series \( \Delta z_{(m)} \). According to the digital signal filtering method, \( \Delta z_{(m)} \) is regarded as discrete sampling signal, which is converted into frequency domain signal by discrete Fourier transform[7], and the frequency domain filtering of height difference series is realized by low-pass filter. The filtering model can be formulated as follows:

\[
F'(k) = H(k)F(k)
\]

\[
(2)
\]

\( F(k) \) is Fourier transform of height difference series, \( H(k) \) is the transform function of low-pass filter[8]. \( F'(k) \) is generated by filtering high frequency information of \( F(k) \). Than, inverse Fourier transform[7] is performed on \( F'(k) \) to obtain \( \tilde{z}_{(m)} \), a new height difference series with uniform variation of data values(figure 2(b)). Substitute to equation(1) to obtain single frame twin point cloud fitting model as follow:

\[
\tilde{z}_{(n)} = \begin{cases} 
  z_{(1)} & n = 1 \\
  z_{(n-1)} + \Delta \tilde{z}_{(n-1)} & n > 1 
\end{cases}
\]

\[
(3)
\]
2.4. Isolated Point Retrieving

The twin point cloud fitting model is obtained by frequency domain filtering contains two messages. Firstly, when the X and Y coordinates of the original isolated points have not changed, the Z coordinate value has been greatly corrected, but it is still unable to determine the specific spatial position of the isolated point. Secondly, frequency domain filtering cause Z coordinate value of non-isolation points distortion in a certain range. Based on this information, the twin point cloud fitting model is compared with the original twin point cloud (figure 3(a)). Threshold comparison using the difference of Z coordinates of points with the same index number in two data sets. Comparative methods can be formulated as follows:

\[
K \geq |z_{(a)} - z_{(a)}| \tag{4}
\]

If the Z coordinate difference between two points, have common index number i, is less than the threshold K, the point i in the original twin point cloud frame is considered to be a non-isolation point. Otherwise, it is determined as an isolated point. We can obtain the twin data height set \(z_{(p)}\) when all isolated points height information in \(z_{(a)}\) are filtered out. Twin data frame \(\hat{P}_{(i)} = (x, y_{(p)}, z_{(p)})\) is constructed (figure 3(b)). \(y_{(p)}\) is a set of Y coordinate value corresponding to \(z_{(p)}\) in \(P_{(i)}\), and the X coordinate value is fixed in the single-frame twin point cloud. The target twin data is obtained when original twin point cloud filter all isolated points.

![Figure 2. Algorithm Diagram 1.](image)

3. Algorithm Example and Analysis

3.1. Experimental Results

In order to validate that the algorithm can satisfy the isolated data filtering of the original twin point cloud obtained in the complex storage space environment, the measured original twin point cloud of a cement factory in Zhejiang Province are selected for experimentation. Figure 4(a) shows the original twin point cloud of the sand mound slope obscured by the beam of the factory building. Figure 4(b) is a single-frame twin point cloud fitting model. It is obvious that under the premise of small distortion of non-isolation points in the fitting model obtained by algorithm steps 1 and 2, the Z coordinate value...
of isolated point is greatly corrected.

![Figure 4. Twin point cloud 1.](image)

Figure 4. Twin point cloud 1.

Figure 4(c) shows the process of isolated point position retrieving in the original twin point cloud according to the twin point cloud fitting model. By setting an appropriate threshold, the isolated points in the original twin point cloud can be completely filtered, while the characteristics of non-isolation points are still maintained. Figure 4(d) shows the target twin data.

We choose another set of measured original point cloud for the experiment, as shown in Figure 5(a). Twin point cloud fitting model (figure 5(b)) is obtained when the algorithm completes process from step1 to step2. Finally, the intact twin data are obtained as shown in figure 5(c).

![Figure 5. Twin point cloud 2.](image)

Figure 5. Twin point cloud 2.

3.2. Algorithm Efficiency Evaluation

To validate the efficiency of the proposed algorithm, according to the filtering accuracy evaluation formula proposed in literature [9]:

\[ P_1 = \frac{N_{TP}}{N_{TP} + N_{FP}} \]  

\( N_{TP} \) is the number of non-noise points are correctly judged as non-noise points. \( N_{FP} \) is the number of noise points mistaken for non-noise points. According to the filtering total error formula proposed in literature [10]:

\[ P_2 = \frac{N_{TP} + N_{FP}}{N} \]
\( N_{TP}' \) is the number of non-noise points mistaken for noise points, which had been filtered in the Step3 of algorithm, \( N \) is the total of points in the original twin cloud. Statistics of experimental results are shown in table 1:

| Parameter | Data set 1 | Data set 2 | Data set 3 |
|-----------|------------|------------|------------|
| \( N \)   | 3615       | 3176       | 8197       |
| \( N_{TP} \) | 3363       | 2937       | 7891       |
| \( N_{TP}' \) | 12         | 35         | 260        |
| \( N_{FP} \) | 0          | 0          | 0          |
| \( P_1 \) | 1          | 1          | 1          |
| \( P_2 \) | 0.0033     | 0.0110     | 0.0317     |

The results, three experiment of isolated points filtering, show the algorithm can remove all isolated points in the original twin point cloud, and the filtering accuracy \( P_1 \) reaches 100%. Under this premise, the algorithm keep characteristics of original twin point cloud as far as possible, the rate of filtering total error is less than 0.05. Therefore the veracity of twin data is greatly guaranteed.

There are two reasons why non-isolation twin points are filtered. Firstly, if the amount of twin points in the single twin point cloud frame is even(amount of height difference data is odd), the algorithm will remove the last point when discrete Fourier transform is performed(figure 6(a) and (b) show the No.10 point is removed from twin point frame). Secondly, when the threshold \( K \) is setting small, few non-isolation twin points could be mistakenly removed in the process of Isolated point retrieving(figure 6(c)). The filtering result with small error as show in figure 6(d).

4. Conclusion
In order to solve the problem of twin data acquisition during the establishment of digital twin stereoscopic warehouse, a high-precision twin point cloud filtering algorithm is proposed in this paper. All the isolated points in the original twin point cloud can be filtered by constructing twin point cloud fitting model. It provides a reliable support for the virtual models and information expansion in the digital twin stereoscopic warehouse system. According to the favorable experiment results, the fast-filtering algorithm has the characteristics of simple implementation and accurate isolated points filtering. It also provides new research ideas for point cloud filtering.
References
[1] Song Y, Shi Y W, Yu J S, Tang D Y, Tao F. 2019 J. Computer Integrated Manufacturing Systems 25 1559-67.
[2] Grieves M and Vickers J. 2017 Digital twin: mitigating unpredictable, undesirable emergent behavior in complex system [J] Springer International Publishing.
[3] Gu J X, Yang B S, Dong Z and Yang C H. 2020 J Bulletin of Surveying and Mapping 6 134-140.
[4] Tao F et al. 2019 J Computer integrated manufacturing systems 25 1-18.
[5] Huang Z W, Liu F, Hu G W 2017 J Acta Optica Sinica 37 353-362.
[6] Liu B et al 2015 J Laser & Infrared 45 117-122.
[7] Zhou L Q and Su F 2005 M Fundamentals of Digital Signal Processing.
[8] Wang D W, Jia R C and Wang H Y 2012 J Modern Electronics Technique 35 71-72.
[9] Zhang J M, Chen F J and Long J L 2021 J Laser & Optoelectronics Progress 58 0610015.
[10] Huang S Y, Liu L M, Dong J and Fu X J 2020 J Opto-Electronic Engineering 47 190688.