Track Data Split Measure Based on Least Square Algorithm

Xingfu Zhang, Yanbo Li, Lei Liu, Feng Zhao, Jingfeng Song
College of Computer Science and Technology
Heilongjiang Institute of Technology
Harbin, China
portzhang@foxmail.com

Abstract—**This paper describes the split measure algorithm about tracks data. In accordance with the needs of the project, the algorithm can be used to measure the two parts, the three parts and the five parts of the data and then combine them together. The algorithm is based on the overlap area of each measurement. And the best stitching position can be obtained by using the least square method. Error analysis of two parts, three parts and five parts measure respectively is carried out. For several measure requirements, Error analysis shows that the theoretical maximum error of the algorithm is less than the actual requirements of the project. Finally, we selected 300 data from a sine wave, and using these 300 data to simulate the 3 meters long track. Then to simulate five measurements, and every two measurements overlap 0.5 meters. Experimental data analysis shows that the algorithm can indeed meet the requirements of practical application.**

Keywords—data splice; data fitting; least square algorithm

I. INTRODUCTION

Engineering needs to measure the straightness of the railway tracks. Straightness measurement is an important means to verify the eligibility of the tracks, so many scholars engaged in research in this field [1-2]. Due to the limited length of the measuring tool, it is necessary to measure the different parts of the tracks respectively, and the two measurements have a certain area overlap. How to splice the multiple measurements properly is a problem that needs to be attached to the application of engineering. The accuracy of the splicing is directly affected by the accuracy of the final measurement. In this paper, a splicing algorithm is used to guarantee the multiple measurement data in a certain error range.

II. ALGORITHM DESCRIPTION

Assuming that the data are measured on the track L, and a total of two measurements are measured. The measurement results were recorded as $L1$ and $L2$. The length of each measurement is 1 meters. $L1$ and $L2$ should coincide with 0.5 meters in theory. As shown in Figure 1.
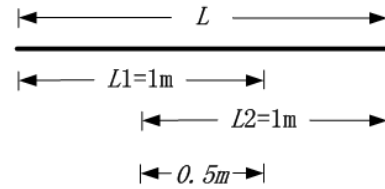
Fig. 1. Two measurements on the tracks

In the actual measurement process, we can not guarantee that the coincidence is just 0.5 meters. Therefore, there is an error in measurement process. Assuming the error is $\pm l$ millimeters, and measurement n data per millimeter.

A. Data preprocessing

$1000 \times n$ data were measured respectively on $L1$ and $L2$ under ideal conditions. In the second half of $L1$ in order to take $(500+l) \times n$ data, and denoted as $L1' = \{x_i, i = 1, 2, ..., (500+l)n\}$. Then in the first half of $L2$ in order to take $(500+l) \times n$ data, and denoted as $L2' = \{y_j, j = 1, 2, ..., (500+l)n\}$.

Two measurements on the $L1$ and $L2$ may be different on baseline, so we need to standardize the measurement data. First, the mean value of measurement data $L1$ is calculated. Represented by $\bar{x}$, and $\bar{x} = \sum_{i=1}^{(500+l)n} x_i \Big/ ((500+l)n)$. And the mean value of measurement data $L2$ is calculated. Represented by $\bar{y}$, and $\bar{y} = \sum_{j=1}^{(500+l)n} y_j \Big/ ((500+l)n)$. Finally, we use the following formula to standard data.

$$L1'' = \{m_i = x_i - \bar{x}, i = 1, 2, ..., (500+l)n\} \tag{1}$$

$$L2'' = \{n_j = y_j - \bar{y}, j = 1, 2, ..., (500+l)n\} \tag{2}$$

B. Segmented data connection

To find the most similar position to connect. Using Euclidean distance as the similarity measure, the smaller the

Euclidean distance is, the greater the similarity is. The implementation of the pseudo code is as follow.

```
for(z=1;z<=2 ln ;z++)
{
  if（z>1）
  {
    L1" = L1" −{m1} ;
    L2" = L2" −{n(500+l)n−z+1} ;
  }
  a[z] = √( Σ(i=1)^((500+l)n−z+1) (mi − ni)^2 ) ;
}
result = min( a[z] );
```

The Z value is obtained when the minimum value of 'result' is obtained. Set the minimum value for Z_{min}. In the end, we can determine the part of the coincidence is $\{m_{Z_{min}},...,m_{(500+l)n}\}$ in $L1"$ and $\{n_1,...,n_{(500+l)n−Z_{min}+1}\}$ in $L2"$.

For coincidence data, the final data is the mean of the corresponding data of $L1$ and $L2$. Then after the connection of the 1.5 meters data is as follow.

$$L = \{ m1, m2, ..., m_{Z_{min}} −1,$$
$$\left(m_{Z_{min}} + n_1\right)/2, ..., \left(m_{(500+l)n} + n_{(500+l)n−Z_{min}+1}\right)/2, \quad (3)$$
$$n_{(500+l)n−Z_{min}+2}, n_{(500+l)n−Z_{min}+2}, ..., n_{(500+l)n} \}$$

C. Calculate the maximum difference

Linear fitting with least square method [3-5]. To facilitate this section description, we redefine the formula 3 in the data symbols as follow.

$$L = \{y_1, y_2, ..., y_{150n}\} \quad (4)$$

We define the order number and L as dot pairs as follow.

$$L(x, y) = \{(1, y_1), (2, y_2), ..., (150n, y_{150n})\} \quad (5)$$

And linear fitting with the data. The equation is as follow.

$$Y = kX + b \quad (6)$$

Use formulate 7 and 8 to solve the value k and b.

$$\sum_{i=0}^{150n} (y_i − b − kx_i) = 0 \quad (7)$$

$$\sum_{i=0}^{150n} x_i (y_i − b − kx_i) = 0 \quad (8)$$

Calculate the distance from the point to the line. Calculate the maximum difference, to get any point to the distance of the fitting line d. $d = |kx_i − y_i + b| / \sqrt{(k^2 + 1)}$. Finally, we can get the maximum difference $\max(d)$.

III. ERROR ANALYSIS

In the actual measurement, the length of each measurement is 1 meters, and the data collected is 1000. Allowable error is between -4 millimeters and +4 millimeters. And the ideal coincidence distance is 500 millimeters. The actual measurements allow the error to 110 millimeters to +10 millimeters. Then the two segments of the data is matched for circulation. Using the least square method, the maximum error is 8/1000=0.008 millimeters. To correct the two measurement data and the correction error is 50%, then a correction procedure produces a 0.001 millimeters error. From this we can get the following results.

When the two measuring distance is 1.5 meters, the maximum error is 0.008*2*50% + 0.001 = 0.009 millimeters. When the three measuring distance is 2 meters, the maximum error is 0.008*3*50% + 0.001*2 = 0.014 millimeters. When the five measuring distance is 3 meters, the maximum error is 0.008*5*50% + 0.001*4 = 0.024 millimeters.

IV. EXPERIMENTAL ANALYSIS

In the following experiment, we collect the standard data from the sine wave. 3 meters data were obtained by the 5 measurement. We use S1, S2, S3, S4, S5, respectively, to express the results of the five measurements. The coincidence degree of each measurement is 0.5 meters. The length of each measurement is 1 meters. Rotate θ degrees of the standard data as stochastic noise. $−\arctan(4/1000) \leq \theta \leq \arctan(4/1000)$.

The direction of the S1 is the standard direction, so we don't add noise to S1. In addition to S1, the other four segments are added to random noise.

Set a sine wave and the period is 1 meters. Then the ten data acquisition and splicing experiment are carried out. The following table is one of the experimental results. The maximum standard data value is 4 and the minimum standard data value is -4.

TABLE I.   EXPERIMENTAL 1

unit of measurement: millimeter

data	stochastic noise (K)	adding noise data		corrected data	
		maximum	*minimum*	*maximum*	*minimum*
S1	-	-	-	-	-
S2	8.6173e-05	3.9354	-4.0216	4	-4
S3	0.002541	3.3852	-5.9262	4	-4.0001
S4	0.0023587	2.2486	-4.6073	3.9997	-4.0001
S5	0.0011545	3.7156	-4.8701	3.9999	-4.0003

Set a sine wave and the period is 0.5 meters. Then the ten data acquisition and splicing experiment are carried out. The following table is one of the experimental results.

TABLE II.   EXPERIMENTAL 2

unit of measurement: millimeter

data	stochastic noise (K)	adding noise data		corrected data	
		maximum	*minimum*	*maximum*	*minimum*
S1	-	-	-	-	-
S2	-0.0019001	5.1904	-3.2903	4.0002	-3.9999
S3	0.0024081	3.7036	-6.1117	4	-4.0001
S4	-0.0037662	6.3651	-2.5989	4.0003	-3.9998
S5	0.0034308	3.5804	-7.0112	4	-4

Set a sine wave and the period is 2 meters. Then the ten data acquisition and splicing experiment are carried out. The following table is one of the experimental results.

TABLE III. EXPERIMENTAL 3

unit of measurement: millimeter

data	stochastic noise (K)	adding noise data		corrected data	
		maximum	*minimum*	*maximum*	*minimum*
S1	-	-	-	-	-
S2	-0.0037706	4.1815	-0.41082	4	-4
S3	-8.0789e-05	0.068142	-3.9597	4	-4
S4	-0.0026566	6.6539	-4	4	-4
S5	0.0038294	2.2725	-3.813	4	-4

From the above data we can see, the experiment of 1000 data is collected on a sine wave with a period of 0.5 meters, once connected experiment's maximum error is 0.0014 millimeters and the minimum error is 0.0006 millimeters. Five times connection experiment's maximum error is 0.0008 millimeters and the minimum error is 0.0002 millimeters. The experiment of 1000 data is collected on a sine wave with a period of 1 meter, once connected experiment's maximum error is 0.0004 millimeters and the minimum error is 0.0004 millimeters. Five times connection experiment's maximum error is 0.0008 millimeters and the minimum error is 0.0002 millimeters. The experiment of 1000 data is collected on a sine wave with a period of 2 meter, once connected experiment's maximum error is 0 millimeters and the minimum error is 0 millimeters. Five times connection experiment's maximum error is 0 millimeters and the minimum error is 0 millimeters. We can draw the following conclusions. The overall trend is that the smaller the data fluctuations, the more accurate the reconstructed data. All the experiments meets the requirement of the engineering application of the single connection error less than 0.01 millimeters, and the 5 time error is less than 0.04 millimeters.

V. CONCLUSION

This paper mainly summarizes the classical corner detection algorithm, and then the recently improved algorithms and its application are introduced. Finally, the paper points out that the corner detection is still a research hot spot in the future.

REFERENCES

[1] H Tanaka, K Tozawa, H Sato, M O-Hori, H Sekiguchi, "Application of a New Straightness Measurement Method to Large Machine Tool, " CIRP Annals - Manufacturing Technology, vol 30,1981, pp.455-159.

[2] HJ Pahk, JS Park, I Yeo, "Development of straightness measurement technique using the profile matching method, "International Journal of Machine Tools & Manufacture, vol 37,1997, pp.135–147.

[3] O Axelsson, "A generalized conjugate gradient, least square method," Numerische Mathematik, vol 51,1987, pp.209–227.

[4] V Radeka, N Karlovac, "Least-square-error amplitude measurement of pulse signals in presence of noise," Nuclear Instruments & Methods, vol 52,1967, pp.86–92.

[5] C Latry, V Despringre, C Valorge, "Automatic MTF measurement through a least square method, " Proceedings of SPIE, 2004, 5570:233-244.