RESEARCH ARTICLE OPEN ACCESS

Preliminary Research on Data Abnormality Diagnosis Methods of Spacecraft Precision Measurement

Rui Zhang[1,2], Huanhuan Chen[1], Guangyun Li[2],Zaihua Yang[3]

[1]School of Software, North China University of Water Resources and Electric Power, Zhengzhou 450045, China;
[2]School of Navigation and Aerospace Engineering, Information Engineering University, Zhengzhou 450011, China;
[3]Beijing Institute of spacecraft Environment Engineering, Beijing, 100094, China)

Abstract
For precision measuring the satellite equipments, providing technical support for satellite assembly, combined with satellite small size, complex structure, satellite equipment shapes vary, and other characteristics, presently, indirect method that using electronic theodolite to measure cube mirror are commonly used to obtain the relative attitude of the respective devices. But in the actual measurement process, there are measurement errors in the measurement data. How to detect anomalies in the data is the focus of this study. This paper proposes two methods to detect abnormal data, that is mathematical geometric method and outlier detection methods. This paper analyzes their theoretical basis and verifies the feasibility of the two methods through part of the actual measurement data to.
Key Words: precision measurement; abnormality diagnosis; Cubic-Mirror; direction cosine; outlier detection;

I. Introduction

The precision measurement refers to the measurement of the accuracy of satellite equipment installation and the housing structure. In different states of the satellite assembly and testing, we need to test the accuracy of the satellite mounting angle and position repeatedly to detect assembly problems or technical design issues early[1].

To achieve this effect, there is a precondition. It is necessary to achieve the desired measurement data precision. However, in the actual measurement process, due to equipment, environment, surveyors and other factors, the measurement data has errors, that is observing abnormal. Measurement methods currently used mostly electronic theodolite. When the measurement data is normal, this method can be used to detect the satellite equipment installation accuracy, to provide technical support for satellite assembly. While the measurement data is abnormal, this measurement data can’t be used to detect the accuracy of satellite equipment and installation, this measurement data should be discarded and then re-measure.

The key question is how to diagnose abnormal measurement data, to detect the correctness and validity of the measurement data, which can ensure individual components attached to the set position accurately, and can readily detect changes in the relationship between the various components in order to ensure the stability of the whole system and achieve the desired effectiveness of the work[2].

II. measuring methods

With the development of modern science and technology, the traditional measurement instruments have been unable to meet the actual needs of production, urgently needing to develop measurement equipments of high precision, easily using and suitable the shape and position errors in the production fields . Collimation technology is one of the important research areas of measurement tests. Particularly for small size satellites with complex structures, and characteristic points of the device can't be measured directly using total station. We can only obtain the relative posture of each device by cube mirrors indirectly. Specifically, which use theodolite to aim cube mirror, constantly adjusting the theodolite position and posture to make collimation axis direction consistent with the normal direction of the cubic mirror.

In China's industrial sectors, many large equipments (such as satellites, spacecraft, etc.) installation and adjustment must be defined strictly coordinate and ensure high-precision electronic theodolite and other equipment ready to build and restore the device coordinates. This paper selected cubic mirror coordinates as measurement coordinates.

2.1 Definition of Cubic Mirror Coordinate

Cubic mirror is composed of six mutually orthogonal mirrors, normal of six surfaces perpendicular or parallel to each other. In this paper, the geometric center of the mirror cube is chosen as the origin, then we optionally choose two mutually perpendicular plane to establish the cubic mirror coordinate system $O-xyz$. As shown in Figure 1, the geometric center of the cube mirror as the origin O, positive normal is z-axis, the normal of right side is y-axis, right hand criterion to determine the positive x-axis.

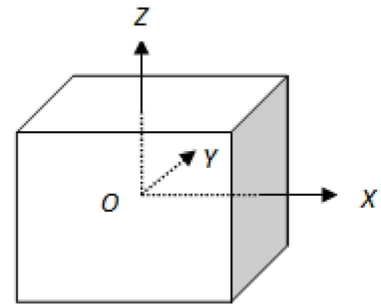

Fig 1. Definition of Cubic Mirror Coordinate

2.2 Establishment of cubic mirror coordinates

With two theodolites T1 and T2, we measure the cubic mirror by the use of collimation to determine the coordinate axes of cubic mirror coordinates, as shown in Figure 2.

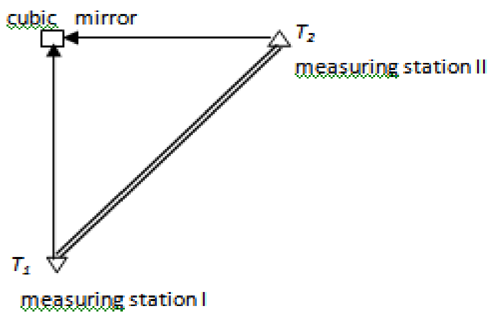

Fig 2. Collimating Measuring of Cube Mirror

Supposing the observed value of two instruments were α_i and $v_i (i=1,2)$ .Cubic mirror collimation has observables α_1 , α_2 , v_1 , v_2 . There are observables α_{12} and α_{21} when the two theodolites mutually aim. We establish measurement coordinate system, the origin of the coordinates is the triaxial center of first instrument, Z-axis is the perpendicular axial direction of the first instrument, Up is positive. The projection on the horizontal plane of the first instrument, when it aimed at the second instrument, is the direction of the X axis, the forward pointed to the second instrument. Y-axis is determined by the right-hand rule. In the measurement coordinate system, for both the collimation direction, $z_1(i_1, j_1, k_1)$ and $z_2(i_2, j_2, k_2)$ vectors have the following formulas:

$$i_1 = \cos v_1 \times \cos \beta_1$$
$$j_1 = \cos v_1 \times \sin \beta_1$$
$$k_1 = \sin v_1 \qquad \text{(Formula1)}$$
$$i_2 = \cos v_2 \times \cos \beta_2$$
$$j_2 = \cos v_2 \times \sin \beta_2$$
$$k_2 = \sin v_2$$

Among them,

$$\beta_1 = \alpha_{12} - \alpha_1, \beta_2 = 180^0 - (\alpha_2 - \alpha_{21}).$$

If the vectors z1 and z2 is vertical in the three-dimensional space, that is,

$$z_1 \times z_2 = \sin v_1 \sin v_2 + \cos v_1 \cos v_2 \cos(\beta_2 - \beta_1) = 0,$$

then, these two vectors can be seen as the two axes of a space rectangular coordinate system, then, we can get the third axis through cross product operation bteween the vector Z1 and Z2. So far, the cube mirror coordinates is established. By using the direction cosine law, transformation matrix that transform from the cube mirror coordinates to measurement coordinates is R_1 , As shown in Formula2.

$$R_1 = \begin{bmatrix} \cos\alpha_x & \cos\alpha_y & \cos\alpha_z \\ \cos\beta_x & \cos\beta_y & \cos\beta_z \\ \cos\gamma_x & \cos\gamma_y & \cos\gamma_z \end{bmatrix} \text{(Formula2)}$$

Among them, α_x is the angle between x-axis of the cubic mirror coordinates and X-axis of the measurement coordinates, β_x is the angle between x-axis of the cubic mirror coordinates and Y-axis of the measurement coordinates, and so on[3].

III. Data abnormality diagnosis method

3.1 Math geometry method

According to the above derivation and combining the nature of vector cross products, we can represent the three vectors by the coordinates, As shown in Figure 3.

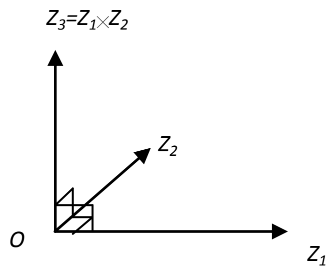

Fig 3. Vector Relationship Diagram

$z_1 = (a_x, a_y, a_z), z_2 = (b_x, b_y, b_z)$, and i, j, k represent the unit vectors of axes z1, z2, z3. Z3 is the result vector cross product operation between z1 and z2, that is,

$$\begin{aligned} z_3 &= z_1 \times z_2 \\ &= (a_x i + a_y j + a_z k) \times (b_x i + b_y j + b_z k) \\ &= a_x b_x (i \times i) + a_x b_y (i \times j) + a_x b_z (i \times k) + \\ &\quad a_y b_x (j \times i) + a_y b_y (j \times j) + a_y b_z (j \times k) + \\ &\quad a_z b_x (k \times i) + a_z b_y (k \times j) + a_z b_z (k \times k) \end{aligned}$$

The cross multiplication rule of unit coordinate vector is:

$$i \times i = j \times j = k \times k = 0$$
$$i \times j = k, j \times k = i, k \times i = j$$
$$j \times i = -k, k \times j = -i, i \times k = -j$$

Following

$$z_3 = (a_y b_z - a_z b_y)i + (a_z b_x - a_x b_z)j + (a_x b_y - a_y b_x)k \qquad \text{(Formula 3)}$$

According to the definition of the direction cosine in space analytic geometry, the vector $(\cos\alpha, \cos\beta, \cos\gamma)$, whose coordinate is the direction cosine of vector r, is the unit vector e_r .

With the rotation matrix R_1, there are:

$$\cos^2\alpha_x + \cos^2\beta_x + \cos^2\gamma_x = 1$$
$$\cos^2\alpha_x + \cos^2\alpha_y + \cos^2\alpha_z = 1 \qquad \text{(Formula 4)}$$

The rest rows and column all meet the nature.

According to formula 3 and formula 4, we can initially judge whether there are abnormal measurement or not. In decimal form, measurement data error allowable range is: $\pm 10^{-4}$.Table 1 shows the rotational component of cubic mirror(x,y,z) in the measurement coordinate system(X,Y,Z).

Table 1. Measurement Data Sample

	X	Y	Z
x	30.0426	64.6591	74.9480
y	120.0379	42.6544	62.8274
z	89.7436	121.6896	31.6908

Verification using Equation 4, the results are as follows:

$$\cos^2 30.0426 + \cos^2 64.6591 + \cos^2 74.9480 = 0.999984$$

$$\cos^2 89.7436 + \cos^2 121.6896 + \cos^2 31.6908 = 1.000001$$

$$\cos^2 30.0426 + \cos^2 120.0379 + \cos^2 89.7436 = 0.999949$$

Measurement data error are within the allowable range. It indicates that there is no abnormality in the set of data.

3.2 Outlier Detection Methods

Outlier detection (also known as anomaly detection) is the process of finding data objects with behaviors that are very different from expectation. Assuming that a given statistical process is used to generate a set of data objects. An outlier is a data object that deviates significantly from the rest of the objects, as if it were generated by a different mechanism. For ease of presentation within this paper, we may refer to data objects that are not outliers as “normal” or expected data. Similarly, we may refer to outliers as “abnormal” data [4].According to the principle of outlier detection methods, outlier detection methods can be used to find abnormality in spacecraft accurate measurement data.

Outlier detection methods make assumptions about outliers versus the rest of the data. According to the assumptions made, we can categorize outlier detection methods into three types: statistical methods, proximity-based methods, and clustering-based methods. Here clustering-based methods were chosen.

The notion of outliers is highly related to that of clusters. Clustering-based approaches detect outliers by examining the relationship between objects and clusters. Intuitively, an outlier is an object that belongs to a small and remote cluster, or does not belong to any cluster. For outlier object that does not belong to any cluster, it can be seen as a small cluster which contains only one object. This makes it possible to draw general method for detecting outliers, on behalf of such methods is FindCBLOF algorithm [5].First, FindCBLOF algorithm uses Squeezer algorithm to partition data objects in the data set into different clusters, then, based on the number of data objects in different clusters, these clusters sort descending, then, based on user-defined parameters α, these clusters is divided into large clusters and small clusters. Each data points assigned a *cluster-based local outlier factor* (CBLOF). For a point belonging to a large cluster, its CBLOF value is the product of the cluster’s size and the similarity between the point and the cluster. For a point belonging to a small cluster, its CBLOF is calculated as the product of the size of the small cluster and the similarity between the point and the closest large cluster.

CBLOF defines the similarity between a point and a cluster in a statistical way that represents the probability that the point belongs to the cluster. The larger the value, the more similar the point and the cluster are. The CBLOF score can detect outlier points that are far from any clusters. In addition, small clusters that are far from any large cluster are considered to consist of outliers. The points with the lowest CBLOF scores are suspected outliers.

Part of aerospace measurement data were selected as the experimental samples. As shown in Table 2. Known that No. 51, 52 tuple are abnormal data, and they form a small cluster. It is suitable to use FindCBLOF algorithm for handling this situation. But Squeezer algorithm used in FindCBLOF algorithm applies only to deal with nominal attributes, not for numeric attributes and ordinal attributes, but in this table clolumn2 to 5 are nominal property, 6 to 11 are numerical property, so, it is not suitable to adopt Squeezer algorithm for Clustering. Therefore, it is need to transform Squeezer algorithm or design a new clustering algorithm.

Table 2. Aerospace measurement data samples

No.	**Spacecraft Model**	**Benchmark cubic mirror**	**Measuring cubic mirror**	**Measurement mode**	**XX**	**XY**	**XZ**	**YX**	**YY**	**YZ**
45	Shenzhou VI	Base cubic mirror	Star Min2	BDWOPD	0.5243682913	1.1284916716	1.3080816883	2.0951318059	0.7445204579	1.0965493538
46	Shenzhou VI	Base cubic mirror	Star Min2	BDWOPD	0.5243562485	1.1285037144	1.3080837827	2.0951023098	0.7444953252	1.0965488302
47	Shenzhou VI	Base cubic mirror	Star Min2	BDWOPD	0.5250550783	1.1153722062	1.8135543470	2.0959233127	0.7120424985	2.0031911770
48	Shenzhou VI	Base cubic mirror	Star Min1	BDWOPD	0.5250461771	1.1153823291	1.8135548706	2.0959430349	0.7120597773	2.0031910025
49	Shenzhou VI	Base cubic mirror	Star Min1	BDWOPD	0.5250025439	1.1154327691	1.8135590594	2.0959294213	0.7120740890	2.0032250364
50	Shenzhou VI	Base cubic mirror	Star Min1	BDWOPD	0.5250346580	1.1153957681	1.8135560923	2.0959318648	0.7120587301	2.0032023471
51	Shenzhou VI	Base cubic mirror	Cm3	BDWPD	0.1919862177	0.2094395103	0.2268928028	0.1919862177	0.1919862177	0.1919862177
52	Shenzhou VI	aaa	bbb	BDWOPD	0.2094395102	0.2268928028	0.2443460953	0.2617993878	0.2792526803	0.2984513021

In Squeezer algorithm, similarities between any tuple t and cluster C are calculated. The maximum value of these similarities was compared with a given threshold, if greater than the threshold value, the tuple t is added to the cluster C; if less than the threshold, creating a new cluster of a tuple t. Method Squeezer algorithm computing the similarity is only suitable for the treatment nominal attributes. So, we have to improve this method so that it can handle nominal attributes and ordinal attributes. Here, we use a method to calculate dissimilarity for attributes of mixed types.

Suppose that the data set contains p attributes of mixed type. The dissimilarity d(i,j) between objects i and j is defined as:

$$d(\mathrm{i},\mathrm{j}) = \frac{\sum_{f=1}^{p} \delta_{ij}^{(\mathrm{f})} d_{ij}^{(\mathrm{f})}}{\sum_{f=1}^{p} \delta_{ij}^{(\mathrm{f})}} \qquad \text{(Formula 5)}$$

where the indicator $\delta_{ij}^{(\mathrm{f})}$ =0, if either x_{if} or x_{jf} missing(i.e., there is no measurement of attribute f for object I or object j),or $x_{if} = x_{jf}$ =0, and attribute f is asymmetric binary; otherwise, $\delta_{ij}^{(\mathrm{f})}$ =1. The contribution of attribute f to the dissimilarity between i and j (i.e., $d_{ij}^{(f)}$) is computed dependent on its type:

（1） If f is numeric: $d_{ij}^{(f)} = \frac{|x_{if} - x_{jf}|}{\max_h x_{hf} - \min_h x_{hf}}$, where h runs over all non-missing objects for attribute f .

（2） If f is nominal or binary: $d_{ij}^{(f)}$ =0, if $x_{if} = x_{jf}$; otherwise, $d_{ij}^{(f)}$ =1.

（3） If f is ordinal: compute the ranks r_{if} and $z_{if} = \frac{r_{if} - 1}{M_f - 1}$, and treat z_{if} as numeric.

Above method is applied to complicated situation that data tuples contain both nominal attributes, binary attributes, numeric attributes and ordinal attributes, at the same time, it is able to convert numeric attribute values or ordinal attributevalues to a common interval [0,1]. This method can achieve data normalized or so called dimensionless.

Similarity between data objects i and j can be calculated by the following formula:

sim(i,j)=1-d(i,j) (Formula 6)

After obtaining similarity between data object i and j, we can use the following formula to calculate the similarity between tuple t and cluster C:

$$sim(i,C) = \frac{\sum_{j=1}^{n} sim(i,j)}{n}$$ (Formula 7)

where $j \in C$, n is the number of elements in the cluster C.

Having got a new similarity measure formula, we can use the transformed Squeezer algorithm and FindCBLOF algorithm to detect outliers in the aerospace measurement data. Take the data in Table 2 as an example,set the parameter s = 0.8, a = 0.9, b = 5,and use reformed Squeezer algorithm to partition the dataset into two cluster, as shown in table 3.

Table 3. Clustering result

cluster	Collection of ids	Is Large Cluster
C1	{43,44,45,46,47,48,49,50}	true
C2	{51}	false
C3	{52}	false

After performing FindCBLOF algorithm, CBLOF value for each tuple t was showed in Table 4.

Table 4. Results of the FindCBLOFalgorithm

id	Value of CBLOF
43	7.346251151472542
44	7.346268858887518
45	7.346262830386474
46	7.346271680088458
47	7.346228703874704
48	7.346253149065038
49	7.346276858199897
50	7.346265309892446
51	0.18280162572252798
52	0.3472745326068604

As can be seen from the data in Table 4, theCBLOF values of No.51, 52 tuples are much smaller than the others, so, they are considered outliers. This is consistent with the expected result, which verified the feasibility of this approach for anomaly detection.

IV. Conclusion

This paper firstly introduces measuring method for device on spacecraft and method of establishing coordinates of the cubic mirror and then focuses on two methods for detecting anomalies in measurement data. Mathematical geometric method uses nature of vector cross product and direction cosine to detect anomalies in measurement data. This method has the stability and efficiency. Outlier detection method uses the intrinsic link between the data values, through a series of mathematical operations, to distinguish between normal data and abnormal data. This method is able to discover abnormal data which is correct in the format. Two anomaly detection methods complement each other, and mutual authentication. Integrating use of them can improve the accuracy of aerospace measurement data abnormality diagnosis to some extents.

V. Acknowledgements

This paper was funded by spacecraft precision measurement joint Laboratory, which is jointly established by Beijing Institute of Spacecraft Environment Engineering and PLA Information Engineering University. Thank Beijing Institute of Spacecraft Environment Engineering and Information EngineeringUniversity forthe support and help in requirements analysis, system processes, systems design and other aspects.

References

[1] Li Manli, Yi Zhong, Liu Xiao. The database system for spacecraft precision measurement [J]. Spacecraft Environment Engineering. Feb 2010, Vol27No.1.

[2] Guangyun Li, Zongchun Li.The Principles and Applications of Industrial Measuring System[M]. Surveying and Maping Press. Jan,2011.

[3] Zhen Yang. Research and Application of Surveying Technology about Optical Collimation [D].Zheng Zhou: PLA Information Engineering University.2009.

[4] Jiawei Han,Micheline Kamber, Jian Pei. Data Mining: Concepts and Techniques, Third Edition [M].China Machine Press. 2012.7

[5] Zengyou He,Xiaofei Xu, Shengchun Deng. Discovering cluster-based local outliers. Pattern Recognition Letters 24(2003) 1641-1650.

[6] Zengyou He,Xiaofei Xu, Shengchun Deng. Squeezer: An Efficient Algorithm for Clustering Categorical Data[J].Comput. Sci. &Technol , 2002 Vo1.17No.5.