TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 12, No. 10, October 2014, pp. 7274 ~ 7279
DOI: 10.11591/telkomnika.v12i8.5835

Automated Data Processing for Monitoring Based on Median Algorithm

Yan Du*[1], Mowen Xie[1], Xiaoli Yang[2], Qiuqiang Liu[3]
[1]School of Civil and Environmental Engineering, University of Science & Technology Beijing
Beijing 100083, China, Ph./Fax: +86-10-62334098
[2]Institute of GeoSpatial Information for GeoHazard Application, University of Science & Technology Beijing
Beijing 100083, China, Ph./Fax: +86-10-62333268
[3]China Institute of Geo-Environment Monitoring, Beijing 100081, China, Ph./Fax: +86-10-62192856
*Corresponding author, e-mail: mutulei@163.com

Abstract

As introduction of the automation equipment, hardware automation level on reservoir has greatly improved. For the restriction of the software performance and technical personnel, data reorganization far failed to meet the requirements of hardware. Base on the research on domestic and foreign technique, it is concluded that a set of data transformation methods suitable for the automation system of the small and medium-sized dams. Through the median denoising processing and eigenvalue automatic statistical techniques, a large amount of data can be screened and filtered. The method can solve practical engineering problems and meet the need of automation equipment.

***Keywords**: data-processing, median algorithm, denoising processing, safety monitoring*

1. Introduction

With the introduction of automation equipment, automatic data acquisition equipment in reservoir management has achieved a rapid development. Though the automatic acquisition devices still have some problems, such as reliability and stability, it can not be denied these equipments and instruments have greatly reduced the workload and work intensity of monitoring staff. Over years considerable progress on hardware makes the automatic data processing become a major challenge. Compared with the traditional data compilation, at present the main problems of automated data compilation to be solved are:

1) The large amount of data. Due to automation equipment is very convenient in acquisition and the need in engineering, each point is measured more than a dozen times daily and collected every 2-4 hours, which is far greater than the traditional data reorganization acquisition density, resulting in a geometric increase in the amount of data. Moreover, static deformation data is rich, but information is poor [1]. It is an urgent problem that how to systematically identify monitoring data analysis and analyze dam geotechnical conditions safety.

2) System error. Reliability of the automation of data is questioned by many. In recent years the reliability and stability of monitoring instrument has been greatly improved, the sampling accuracy has reached more than 80%, and even that of some equipment has reached 90%, however, during sampling, there are still extreme value caused by equipment, environmental and other reasons [2]. No matter in the data analysis warning or in later data reorganization [3], they will directly lead to a decrease in reliability and credibility of data analysis [4, 5].

Based on the above consideration, according to experience of automation integration and mathematical statistics methods[6-9], this paper focuses on analyzing reorganiztion methods of dams and finding a set of scientific data integration analysis method which verified by the engineering examples.

2. Current Situation and Methods

By investigating and comparing the domestic and foreign data processing, analyzing domestic developing state and future development mode, this paper is to find a simple and

Received February 21, 2014; Revised July 3, 2014; Accepted July 28, 2014

effective data processing method of automatic recognition suitable for engineering and computer application, meeting the practical needs of the reservoir dams.

2.1. Data Reorganization Situation

According to their own national conditions, different countries vary in degree in the use of automation monitoring system, but mainly use the integrated management that have a great influence in the world. The dam monitoring system invented by the French power company monitored more than 150 dams and analyzed with a simple statistical model of the system in order to find out the structural abnormalities. The MIDAS system in Italy has been successfully used for nearly thirty years, and is known for the use of the hybrid model and deterministic model for online monitoring. Centralized management by professionals is a common feature in these systems which is a very good reference for our country to develope data acquisition processing mode.

In recent years, while applicating MIDAS system, Italy also developed the dam safety assessment decision support system (DAMSAFE). North America and other countries also began to use network and the integration in automated data processing management, like CANARY. in US. Through the Internet, CANARY manage the dam power station data reorganization and dam safety monitoring system, which realized the win-win situation of the management, supervision and hoster and achieved good social benefits.

In China, the decision support system and the expert system require complex knowledge engineering, and a lot of manpower and material resources. So in the overall point, this study is still in the initial stag. The dams safety automation monitoring largely depends on the specific experience and technical level of the operators.

In early 1993, Hohai University and Electric Power Bureau in Fujian Province jointly developed the "Fujian Province expert decision system on hydropower dam safety management" which made great progress in remote monitoring, real-time analysis and network design, and then in 1994 Hohai University developed the "Longyangxia dam safety assessment expert system" and in 2002 Nanjing Water Conservancy Hydropower Science Research Institute developed the "dam safety monitoring data analysis and evaluation system".The two have been successfully applied in practical engineering. In 2007, the Isafety automatic monitoring and management platform using the Internet and professionally managing with the cloud platform fills the domestic blank in this field.

With the development of the data monitoring and the coming of big-data ,increasing specialization, integration and network will come to data reorganization and professional data hosting services will be a mainstream. So efficient and reliable algorithm for collaborative work will ensure the coordination among sub-systems and the high efficiency in every single system, and improve the overall function of automatic monitoring system [9]. A new design method needs clear thinking and considers a wide range of design process. From the view of the whole system, it can coordinate the relationship between the whole and parts and among parts. The processing method should be more scientific and effective, so that data processing which based on experience algorithm in the long-term becomes more theoretical, scientific and rational.

2.2. Denoising Methods Based on the Median Theory

The denoising methods of small and mediam-sized dams data are artificial denoising and automatic denoising. Artificial denoising, namely manually compare and review the data error, is widely used in traditional monitoring, but few in the automatic denoising. For requiring a lot of time, it is only suitable for small dams with less measuring points and lower monitoring frequency. Automatic denoising based on automation programming language increasingly becomes the mainstream method, such as threshold denoising and the wavelet denoising[3] which are very common at present.

However, there are still some problems in automatic denoising method:

1) The algorithm are not unified. it is difficult to find certain algorithm to meet all engineerings' requirements.

2) The interference of human factors. Because it is based on the experience of the operator to determine the algorithm and threshold, it maybe not objective.

3) It is difficult to apply. Because some automatic denoising algorithm is very complex and consume more time [10], the pre-processing task becomes the most critical part of the anlysis [11].

The median denoising algorithm (MDA) can effectively solve the above problems. In the automatic monitoring, when the data acquisition is not normal and even extreme values occurred for system errors, MDA can filter them and get real results as median is taken as general representative. Because median is in the middle position in sequence among the variables , it is not affected by extreme values (maxima and minima). And because the measured median value is an actual value, comparing with the average value it has more practical and statistical significance. So considering the great difference in measuring the variable value and even the huge difference among automatic data, MDA is appropriate to calculate for it not only reducing the influence of system error, but also greatly reducing the amount of historical data, in order to make the subsequent tedious data more effectively.

The project requires fast analysis of the safety of dams, and the calculation method is too complex to apply. Compared with wavelet filtering [3, 12] and other automatic denoising method, MDA has several advantages as followed:

1) The simple algorithm. The computer denoising algorithm is more efficient, and it is widely used especially in the age of big data.
2) MDA does not change the measured value. The actual value acquired makes subsequent data analysis more valuable.
3) The unified algorithm. All of these terms are ambiguous so something more precise is needed before attempting to quantify features[7]. The algorithm needs not to set certain threshold and parameter which can avoid the interference of subjective factors and ,therefore, realize the automatic denoising.
4) Advanced in denoising. According to the formula (1), denoising rate can be as high as 99.99% (considering the equipment correct rate as 90% and the number of sample as 9).

$$F = \sum_{i}^{n} [C_n^i \times P^i \times \overline{P}^{(n-i)}] \tag{1}$$

Where: *F* is noise reduction rate, *P* is the equipment correct rate, *n* is the number of the sample, *i* is the correct number in the samples.

Because the data acquisition is easy and convenient, it is feasible to reduce the number of samples for sample accuracy [13]. In the new period of data integration, the most critical factor in analyzed statistics is not the small number of samples traditionally, but the errors in the large amount of data. How to eliminate abnormal data from abnormal measurement or induction failure in automation equipment is very critical to the accuracy and the results of error analysis. MDA can solve it.

However, MDA eliminates half of the sample data. Superficially these data may not used in the later analysis, but the data are still useful actually. Automated system can offer more data than actural requirement [14], so the number of the samples processed still meet the analysis requirements.

2.3. Technical Scheme

Firstly, in order to eliminate the abnormal data in automatic data and system error, we process automatically data by using MDA. Secondly, according to the statistic method based on characteristic value, we get the solution to process the automatic data. In this way we can provide targeted processing method for small and medium dams.

3. Application Cases

3.1. Automatic Data Investigation

In recent years the reliability and stability of monitoring instrument has been greatly improved and the sampling accuracy has reached more than 80%, but there are still extreme value caused by equipment, environmental and other reasons. Figure 1 shows the history line of the automatic data. This is the automatic data history curve of a reinforced meter in practical engineering. Abnormal data occur because of the system or sensor problems. These peaks belong to the invalid data and the errors make the analysis confused, which may lead to wrong conclusion. All of these terms are ambiguous so something more precise is needed before attempting to quantify features. So it is necessary to process the data.

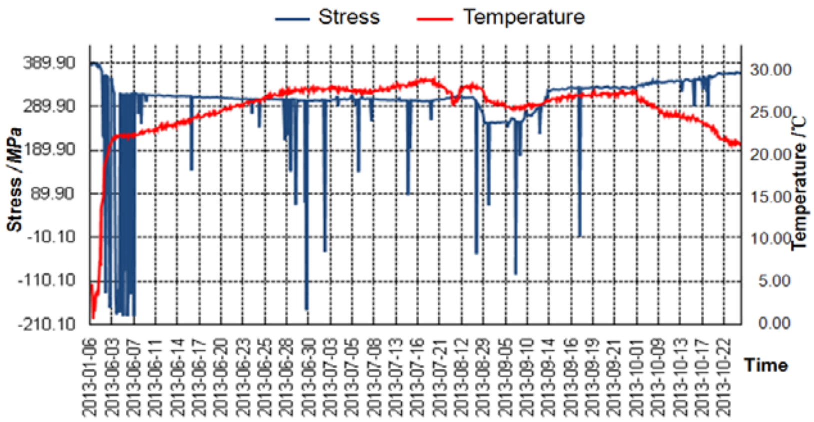

Figure 1. The History Line of the Automatic Data without Data Processing

3.2. The Application of MDA

After using MDA, peaks obviously removed, and the denoising rate was 100%. As shown in fig. 3, the denoising effect is obvious.

The algorithm is as follows:

```
n=1;
    while(n<=fix(length(a)/m))
    for i=1:m
        b(i)=a(i+(n-1)*m);
        c(n)=median(b);
    end
        n=n+1;
end
```

In the context of Gig Data, efficiency is impotent. According to some projects, as we tested, the response time of MDA is within 60 seconds, which is far less than other algorithms. Results are shown in Table 1.

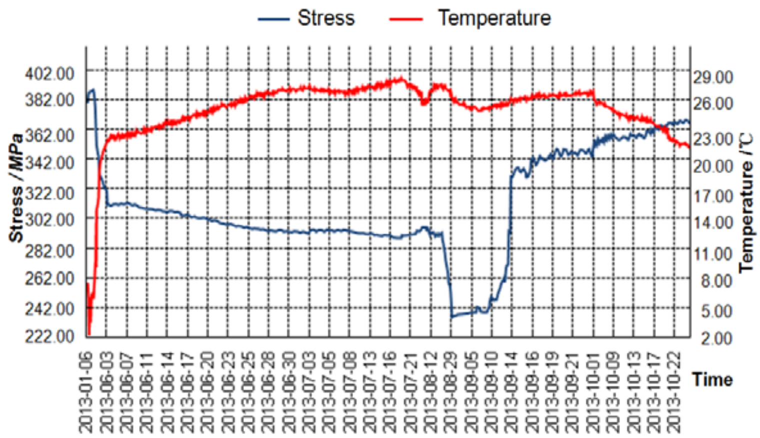

Figure 2. The History Line of the Automatic Data after MDA

Table 1. The Performance of MDA

Program	The number of points	The number of masured values	Denoising time /*s*
Small dam	112	327,040	1.63
Hydropower Station	1014	3,701,100	30.8

3.3. The Application of Automation Algorithm Based on Characteristic Value

Using computer automatically statistical functions [15] can quickly analyze the monitoring data and get the position whose amplitude measures larger. Thus it will greatly reduce the search time for us, provide preliminary reference and focus on the target for data analysis. As shown in Table 2, R205 measuring point of the cross section varies greatly, which in October occurred in the change of 70.63mm. So it requires further attention and analysis for the engineering staff in the practical analysis and monitoring work. The comparison can provide new ideas and references. We can also use the computer information processing function to realize two-dimensional image display.

Table 2. Strain Gauge Observation Results in October Unit：*mm*

Measuring point name	The maximum value	The maximum occurrence time	The minimum value	The minimum occurrence time	Amplitude
R201	9.34	2013/10/19 08:00	4.39	2013/09/20 00:00	4.96
R202	21.72	2013/10/19 08:00	14.78	2013/09/20 00:00	6.94
R203	141.61	2013/10/19 04:00	118.53	2013/09/22 22:00	23.07
R204	264.99	2013/10/19 04:00	243.36	2013/09/23 22:00	21.63
R205	**362.38**	**2013/10/19 06:00**	**291.75**	**2013/10/15 05:00**	**70.63**
R210	-23.01	2013/09/20 22:00	-10.10	2013/10/19 08:00	-12.90
R211	121.24	2013/10/19 04:00	104.52	2013/09/22 22:00	16.72
R212	277.42	2013/10/19 04:00	252.82	2013/09/23 22:00	24.60
R213	169.49	2013/10/19 06:00	151.83	2013/09/21 20:00	17.66
R214	23.79	2013/10/19 08:00	2.16	2013/09/23 22:00	21.63
R215	92.77	2013/10/19 08:00	80.12	2013/09/21 20:00	12.65
R216	45.03	2013/10/19 08:00	32.85	2013/09/20 22:00	12.19

4. Conclusion

Hardware automation equipment in China has become more mature, but the method of processing the automatic data developed slowly which cannot meet the needs of practical engineering of dam. In terms of software, such as data processing and analysis, automatic processing technique has to be further promoted.

Through comparative analysis and engineering practice, the effective combination of median algorithm and eigenvalue statistics, by programming with computer language, can realize initial reorganization function of automatic data . Because of its simplicity in operation, it is suit for small and middle dams automatic data.

The MDA calculates quickly. In the age of big data, this denoising method is simple and general and can quickly provide the required engineering data for analysis.

With the development of monitoring system, data processing will become more and more theoretical, scientific and rational. Specialization, integration and network, managed service will be professional. It will be a mainstream that the dams acquire a professional hosting service.

References

[1] CHING-YUN K, CHIN-HSIUNG L. Monitoring of long-term static deformation data of Fei-Tsui arch dam using artificial neural network-based approaches[J]. *Structural Control and Health Monitoring*, 2013(20): 282-303.
[2] Hu W, YANG X G, ZHOU J W, et al. Gross error denoising method for slope monitoring data at hydropower station[J]. *Telkomnika Indonesian Journal of Electrical Engine*, 2013, 11(10): 5545-5552.
[3] Su H Z, Wu Z R, Wen Z P. Identification model for DAM behavior based on wavelet network[J]. *Computer-AidedCivilandInfrastructureEngineering*, 2007, 22(6): 438-448.
[4] KIBLER K M, TULLOS D D, KONDOLF G M. Learning from dam removal monitoring: challenges to

selecting experimental design and establishing significance of outcomes[J]. *River research and applications*, 2011, 27(6): 967-975.

[5] BUCHHEIT R B, GARRETT J H, MCNEIL S, et al. Automated procedures for improving the accuracy of Sensor-Based monitoring data[C]//*Applications of Advanced Technologies in Transportation*. Boston, 2002: 402-409.

[6] JAYAWARDENA A W, GURUNG A B. Noise reduction and prediction of hydrometeorological time series_ dynamical systems approach vs. stochastic approach[J]. *Journal of Hydrology*, 2000, 228(3/4): 242-264.

[7] RANDOLPH T W, YASUI Y. Multiscale processing of mass spectrometry data[J]. *Biometric*, 2006, 62(6): 589-597.

[8] CHRISTER A H, WANG W. A simple condition monitoring model for a direct monitoring process[J]. *European Journal of Operational Research*, 1995, 82(3): 258-269.

[9] ZHANG D Q, HAN Y B, TANG X Y. Nonlinear_Nongaussian time series prediction based on RBF-HMM-GMM model[J]. *Telkomnika Indonesian Journal of Electrical Engine*, 2012, 10(6): 1214-1226.

[10] Fang B, GUO X Q. Modified allan variance analysis on random errors of MINS[J]. *Telkomnika Indonesian Journal of Electrical Engine*, 2013, 11(3): 1227-1235.

[11] WU C-h, YANG C-h, LO S C, et al. Automatic data mining for telemetry database of computer systems[J]. *Microelectronics reliability*, 2011, 51(2): 263-269.

[12] LIANG G L, XU W Y, TAN X L, et al. DENOISING PROCESSING OF SAFETY MONITORING DATA FOR HIGH ROCK SLOPE BASED ON WAVELET TRANSFORM [J]. *Chinese Journal of Rock Mechanics and Engineering*, 2008, 27(9): 1837-1844.

[13] YAN Du, XIE Mo-wen, MAN Hu. Mechanism of lake area variations and water level changes in NAM CO lake[C]//*Advanced Materials Research*. Shenzhen, 2013: 1685-1692.

[14] ZHAO S H, XIANG T Y, LAI H G. Analysis of compilation optimization of hydrological data in upstream and downstream stations of Three Gorges Project before and after impoundment[J]. *Yangtze River*, 2012, 43(23): 18-22.

[15] SUN E J, ZHANG X K, LI Z X. The Internet of things (IOT) and cloud computing (CC) based tailings dam[J]. *Safety Science*, 2012, 50(4): 811-815.