

## A New Surveillance Method of Machine Status using Big Data

Fuzhen Xie

*School of Mechanical Engineering Xinyu University, Jiangxi Xinyu 338004, China*  
woaikan050913@163.com

### **Abstract**

*As the development of science and technology, especially the network and IT techniques, lots of enterprises are facing the floods of data, which brings new challenges such as the storage, processing, and application under the new era. As in the industry area, the data will increase in tremendous level and their relations are more and more complex. That results in great issues when enterprises are contemplating to using the rich information or knowledge hidden from the huge data sets. In order to investigate the trends of the machine status from big data, this paper examines the coal milling machine's abrasion, which is usually monitored. The surveillance is changing along with the time, which is able to reflect the abrasion trends so that the machine maintenance and breakdown could be easily carried out. Given the characteristics of big data in industrial field, this method proposes a multi-scale system with the entropy and energy to reflect the machine status in large scales. Two experiments are carried out to examine the methods. It is founded that, as the time increasing, the changes are becoming larger.*

**Keywords:** *Surveillance, Big Data, Machine Status, Multi-scale, Entropy, Energy*

### **1. Introduction**

As the development of science and technology, especially the network and IT techniques, lots of enterprises are facing the floods of data. The huge number of data brings new challenges, for example, the storage, processing, and application under the new era. Big Data was proposed to get rid of the challenges caused by the data floods. It was reported that there are four characteristics of the big data: volume, variety, velocity, and variability [1]. As in the industry area, the data will increase in tremendous level and their relations are more and more complex [2]. That results in great issues when enterprises are contemplating to using the rich information or knowledge hidden from the huge data sets.

However, there are some challenges when contemplating to make full use of the big data. The challenges are summarized into 4Vs as follows:

**Volume.** This is the most crucial characteristic of the industrial big data. A complex industrial production process usually generates huge number of data when surveillance points need to be monitored and controlled. For example, a machine with the 1000MW electricity generating set has over 12,464 surveillance points, 1,400 control facilities such as 198 motors, 1,117 electronic switches, and so on. The surveillance system is able to generate immense data for daily usage [3]. It is estimated that the data will be TB level every year, which cause the exactly big data.

**Variety.** The traditional industrial storage system only records the switch information or simulation data. As the advanced materials innovation and development of surveillance methods, CCD image processing devices and voice sensors for surveillance are widely used in extreme complex environments. These devices can capture different types of data like images, videos, voice, and texts, which are usually unstructured data with wide varieties.

**Velocity.** The modern industrial production is large scale with complex architectures. Control circuits and surveillance points are huge. Such devices may generate thousands of MB data within a second. The data should be transferred, processed, and stored with high speed, efficient feedback, and on-time strategy.

**Variation.** The production has multi-layers characteristics, which cause the variations of data from different layers. The dynamics and static statuses of machines, workers, and materials vary a lot, resulting in the variations of the data connection, delays, and complexity. The facilities' defaults will be varied along with the parameters changes.

For the machine status surveillance in industrial field, there are some special characteristics that will bring more challenges for making full use of the data. Firstly, multi-variables dimensional system is common with physical, chemistry, and other principles to combine all the parameters together, forming a complex variable-oriented system. In order to express the process, different parameters must be integrated to build up a high dimensional description, resulting difficulties in understanding the data [4-9]. Secondly, the surveillance data is non-linear. The machine surveillance in industrial area is not linear system. For example, the heating system is a typical non-linear system considering the temperature, compress, and radiation. Such relation causes difficulties in understanding the data. Thirdly, the samples are not distributed normally. The machines are always working in a high efficiency. Take the power supplier for an example, the steel ball coal mill machine used for preparing the coal for generating the power will be utilized in high ratio so that the profit margin will be the optimal. Additionally, considering the abrasion of the mill machine and safety, the system cannot work with a long time of overload. As a result, the peak season and the off-season are able to generate different sample data that is not normal distribution [10].

In order to investigate the trends of the machine status from big data, this paper examines the coal milling machine's abrasion, which is usually monitored. The surveillance is changing along with the time, which is able to reflect the abrasion trends so that the machine maintenance and breakdown could be easily carried out. For this purpose, this paper introduces a new surveillance method of machine status using big data based on the multi-scale system, scale entropy, and scale energy. The mathematic expressions are illustrated for describing the problem and using the method to predict the machine status from big data.

## 2. Multi-scale System

For a nonlinear time variety system  $Y = f(U, t)$ , the input  $U$  and output  $Y$  is carried out under wavelet changes of  $k$  scale.  $u_m \hat{\Gamma} U$ ,  $y_n \hat{\Gamma} Y$  could be divided into the detailed constructed signal  $u_m^1, u_m^2, \dots, u_m^j$ ,  $y_n^1, y_n^2, \dots, y_n^j$ ,  $j = 1, 2, \dots, k$  as well as similar constructed signal  $u_m^a$  and  $y_n^a$ , which are defined as follows:

The detailed and similar constructed signal of  $u_m$  and  $y_n$  ' no  $j$  scale have four energy functions:  $P_{det}(u_m^j), P_{det}(y_n^j), P_{sim}(u_m^a), P_{sim}(y_n^a)$ .

$$P_{det}(u_m^j) = \int u_m^{j2}(t) dt \quad (1)$$

$$P_{det}(y_n^j) = \int y_n^{j2}(t) dt \quad (2)$$

$$P_{sim}(u_m^a) = \int u_m^{a2}(t) dt \quad (3)$$

$$P_{sim}(y_n^a) = \int u_n^{a \lim_{x \rightarrow \infty} 2}(t) dt \quad (4)$$

The total energy of  $u_m$  and  $y_n$  are:

$$P(u_m) = \int u_m^2(t) dt \quad (5)$$

$$P(y_n) = \int y_n^2(t) dt \quad (6)$$

The energy density from the above similar definition, we can get that:

$$Q_{det}(u_m^j) = \frac{P_{det}(u_m^j)}{P(u_m)} \quad (7)$$

$$Q_{det}(y_n^j) = \frac{P_{det}(y_n^j)}{P(y_n)} \quad (8)$$

$$Q_{sim}(u_m) = \frac{P_{sim}(u_m^a)}{P(u_m)} \quad (9)$$

$$Q_{sim}(y_n) = \frac{P_{sim}(y_n^a)}{P(y_n)} \quad (10)$$

According to the definition, the multi-scale of energy matrix could be expressed as  $M_{det}, M_{sim}$ :

$$M_{det} = \begin{bmatrix} 1 & \frac{P_{det}(u_m^1)}{P_{det}(u_m^2)} & \dots & \frac{P_{det}(u_m^1)}{P_{det}(u_m^j)} \\ \frac{P_{det}(u_m^2)}{P_{det}(u_m^1)} & 1 & \dots & \frac{P_{det}(u_m^2)}{P_{det}(u_m^j)} \\ \dots & \dots & \dots & \dots \\ \frac{P_{det}(u_m^j)}{P_{det}(u_m^1)} & \frac{P_{det}(u_m^j)}{P_{det}(u_m^2)} & \dots & 1 \end{bmatrix} \quad (11)$$

$$M_{sim} = \begin{bmatrix} 1 & \frac{P_{det}(y_n^1)}{P_{det}(y_n^2)} & \dots & \frac{P_{det}(y_n^1)}{P_{det}(y_n^j)} \\ \frac{P_{det}(y_n^2)}{P_{det}(y_n^1)} & 1 & \dots & \frac{P_{det}(y_n^2)}{P_{det}(y_n^j)} \\ \dots & \dots & \dots & \dots \\ \frac{P_{det}(y_n^j)}{P_{det}(y_n^1)} & \frac{P_{det}(y_n^j)}{P_{det}(y_n^2)} & \dots & 1 \end{bmatrix} \quad (12)$$

The above matrix presents the input and output variables distribution under different scale with the degree of the energy density. Using the  $M_{det}, M_{sim}$ , the multi-scale system could be defined:

When  $0 < \hat{a}_{ij} < \infty$ , the input of the system has the hefts. The system is called nonlinear time variety system  $Y = f(U, t)$  is a multi-scale of output system. When

$0 < \hat{a} b_{ij} < \infty$  the output of the system has the hefts, the system is a multi-scale of input. While a system has the two characteristics, this system is called input and output multi-scale system [11].

### 3. The Multi-scale Real System

#### 3.1. Characteristics

The input of nonlinear time variety system is not able to measure all the access, if there is a system  $Y = f(V, W, t)$  where  $v_i \hat{\Gamma} V$  is the measurable input,  $w_i \hat{\Gamma} W$  is the non-measurable input.  $w_i$  is a limited length variable that meets:

$$L = \{w_i, \|w_i\|_2 < \infty\} \quad (13)$$

$$\|w_i\|_2 = \int_{-\infty}^{+\infty} |w_i(t)|^2 dt \quad (14)$$

$w_i$  could carry out the Fourier transformation to decompose into different hefts with different frequencies. For a frequency  $F \hat{\Gamma} [f_{jl}, f_{jh}]$ , it meets:

$$P(X_F) = \int_{f_{jh}}^{f_{jl}} X^2(jW)dW \neq 0 \quad (15)$$

Where,  $f_{jl}, f_{jh}$  are the frequency range from the  $k$ -scale wavelet changes of the no.  $j$ .

#### 3.2. Scale Entropy and Energy

For a input and output multi-scale system  $Y = f(U, t)$ , we could carry the  $k$ -scale wavelet change when  $u_i \hat{\Gamma} U$ . Thus,  $u_i$  could be decomposed into  $u_{ij}$ ,  $j = 1, 1, \dots, k, k + 1$ , which is the detailed wavelet reconstruction and  $u_{ik+1}$  is the similar wavelet reconstruction. The scale entropy of  $u_i$ 's no  $j$  scale is  $EP_j(u_{ij}) = -\log q_j$ , where

$$q_j = \frac{P(u_{ij})}{\hat{a} \sum_{m=1}^{k+1} P(u_{im})} \quad (16)$$

The energy of  $u_i$ 's no  $j$  scale is:

$$Y_j(u_{ij}) = \int EP_j(u_{ij}) dt \quad (17)$$

#### 3.3. Problem Description

A complex system needs the surveillance of the machine statuses real-timely. The status parameters may be changing along with the time and it is monotone increasing such as the abrasion. Based on the multi-scale system with the above definition, the problem could be described as follows:

The scale  $k$  and  $u_{ik}$  meet except the limited time interval  $[t_1, t_2]$ ,  $Y_k(u_{ik})$  satisfies

$$Y_k(u_{ik}) = \min(Y_j(u_{ij})) \quad (18)$$

$$\text{and } Y_k(u_{ik}) \square Y_j(u_{ij}), j = 1, 2, \dots, m, j \neq k \quad (19)$$

for the time interval  $[t_1 - Dt, t_1]$  and  $u_{ik}$  meets

$$\int_{t_1}^{t_1+Dt} u_{ik} dt \geq \int_{t_1-Dt}^{t_1} u_{ik} dt \quad (20)$$

when  $t$  changes into  $t + Dt$ :

$$\frac{Dt}{t} \square \frac{u_{ik_{i+1}}}{u_{ik_i}} \quad (21)$$

(20) indicates the monotone non-decreasing of the signal system. (21) presents the slow changing. Thus, the surveillance of the status could be converted into a construct status surveillance function  $R = g(W', Y', t)$  and the mapping logic  $f'$ . Where  $W' \hat{=} W, Y' \hat{=} Y$ .  $R$  or  $R_n$  and  $f'$  meets the multi-scale nonlinear system status parameters reconstruction conditions:

Scale  $m \in k+1$  is the main scale of  $u_i$ . Mapping  $f': u_{im} \rightarrow R$  or  $f': u_{im} \rightarrow R_n$ . " $u_{im_1} \hat{=} u_{im}, u_{im_2} \hat{=} u_{im}, u_{im_1} \neq u_{im_2} \rightarrow f'(u_{im_1}) \neq f'(u_{im_2})$ " exists. Where,  $R_n$  is the wavelet transformation of  $n$  scale reconstruction from  $R = g(W', Y', t)$ .  $u_{im}$  is the no  $m$  scale wavelet transformation of  $u_i$ .

## 4. Multi-scale Status Surveillance based on Big Data

The machine status surveillance method is based on the big data from the industrial information system. These big data could be regarded as input and output multi-scale system signal [12]. Thus, it is possible to establish the status surveillance function which meets the status parameter reconstruction condition. The function could be used for predicting the machine status. This method is based on the standard model and residual multi-scale analysis.

### 4.1 Standard Model

This model uses the object running data, design parameters or analyzed data so as to build up a model that is able to reflect the input and output relationships. The purpose is to build up the corresponding rules among objects' parameters, providing reference information for the status surveillance. During the machine running, the measurable information and the reference information provided by the standard model could be compared to establish the residual sequence [13]. When there is some unstable issues, the residual sequence will be significant changed. The standard model is built up given the input and output. Here, we take the stable standard model (SSM) for example. SSM refers to using the stable machine primary data to build up a model that is able to reflect the parameter and status relation. The standard is considered that in the time interval  $[t_1, t_1 + Dt]$ , the modeling variable  $x_i$  satisfies:

$$d_{x_i} = \frac{x_{i_{\max}} - x_{i_{\min}}}{x_{ie}} \cdot 100\% < d_k \quad (22)$$

where  $x_{i_{\max}}$  and  $x_{i_{\min}}$  are the maximum and minimum value within time interval  $[t_1, t_1 + Dt]$ .  $x_{ie}$  is the stable value,  $d_k$  is variation threshold.

### 4.2. Residual Multi-scale Analysis

The dynamic and stable standard models reflect the relationship of parameters and status under different system situations. Based on the standard models, the running data from

historic machine statuses could be input into the model for getting the prediction of the system changes. The big data warehouse could be input into the following models:

$$y_{i,preDYN} = f_{dym}(x_1, x_2, \dots, x_n) \quad (23)$$

$$y_{i,preSTD} = f_{std}(x'_1, x'_2, \dots, x'_m) \quad (24)$$

where,  $y_{i,preDYN}$  and  $y_{i,preSTD}$  are the dynamic predicted value and stable model prediction value.  $\{x_i | x_i \hat{=} [x_1, x_2, \dots, x_n]\}$  is the dynamic running data.  $\{x'_i | x'_i \hat{=} [x'_1, x'_2, \dots, x'_n]\}$  is the stable running data. Thus, the residual multi-scale from the comparison of both values with real-life data could be obtained by the following formula:

$$\text{The dynamic residual: } err_i^D = y_i - y_{i,preDYN} \quad (25)$$

$$\text{The stable residual: } err_i^{STD} = y'_i - y_{i,preSTD} \quad (26)$$

$y_i, y'_i$  are the real running data and stable running data.

For an input and out multi-scale system, the status surveillance will change slightly due to the slow variations of the status parameters. For a short time interval  $[t_1, t_2]$ , the changes of status parameters are regarded as linear or simple nonlinear like quadrics. Thus, the residual analysis could be carried out by linear or nonlinear functions as well. The covariance could be got by the estimation of dynamic residual wavelet transformation detailed reconstruction:

$$R = cov(W(s, err), W(s, err)) \quad (27)$$

where,  $W(s, err)$  presents the dynamic residual  $err_{DYN}$ 's detailed hefts of the reconstruction in no  $s$  scale. Then, the best estimation from the big data could be calculated from the following formula:

$$\hat{x}_k = A\hat{x}_k + Bu_{k-1} \quad (28)$$

$$P_k^- = AP_k^- A^T + Q \quad (29)$$

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \quad (30)$$

$$\hat{x}_k = \hat{x}_k + K_k(z_k - \hat{x}_k) \quad (31)$$

$$P_k = (I - K_k H)P_k^- \quad (32)$$

## 5. Experiments and Discussions

The experiments come from a real-life power supplier company, which uses the coal for its machines that generate the electricity. To simplify the problem, we use the signal as the example to reflect the machine status. Assume that the big data from the machine running historic data obeys the distribution of following trends:

$$y = \sin(t) + d(t) \quad (33)$$

$$\begin{cases} d(t) = 1, t = 12.56 \\ d(t) = 0, t \neq 12.56 \end{cases} \quad (34)$$

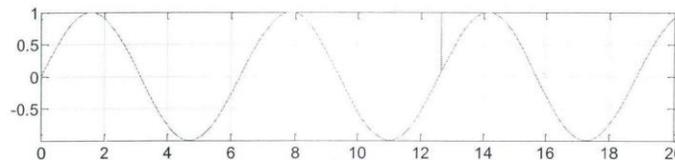
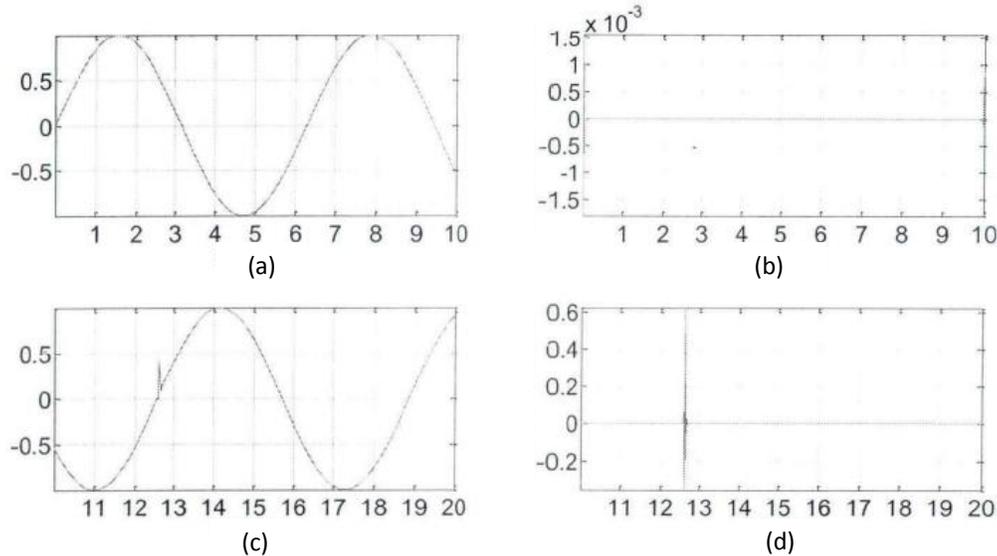


Figure 1. Machine Status Curve

Figure 1 shows the machine status, which is divided into two parts. The first 10 seconds is the first part, which has the signal of  $y = \sin(t)$  to present the machine status. The later 10 seconds is the second part, which has the pulse signal  $d(t)$ . The wavelet transform is carried out for multi-scale analysis. In scale 1, the detailed hefts reconstruction and similar hefts reconstruction are obtained, showing in the following figure 2, which presents the multi-scale analysis of the machine status.



**Figure 2. Multi-scale Analysis of Machine Status (Signal)**

Figure 2 (a) shows the first 10 seconds similar heft reconstruction. There are two peaks and one trough so that the machine will be usually busy with doing the jobs. Figure 2 (b) shows the first 10 seconds detail reconstruction, which is a point that is  $-0.5 \times 10^{-3}$ . It reveals that the machine is in a heavy load. Therefore, the maintenance should be carried out frequently. Figure 2 (c) and (d) show the later 10 seconds status of similar and detailed hefts reconstruction. It is observed that there are two troughs and the machine has wide range of detailed hefts reconstruction. The machine status is occasionally busy and idle.

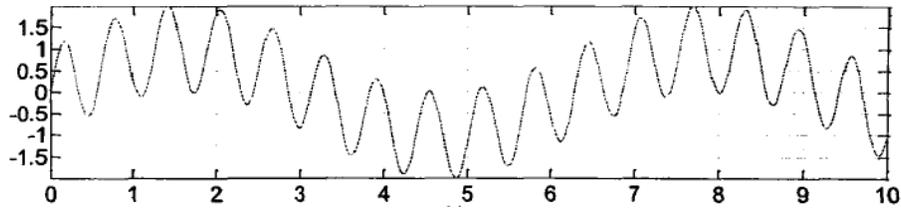
**Table 1. Multi-scale Entropy and Energy of Detailed Hefts Reconstruction in Scale 1**

	1-10s	10-20s
Energy Intensity	1.45e-008	0.001
Multi-scale Entropy	18.04	6.777

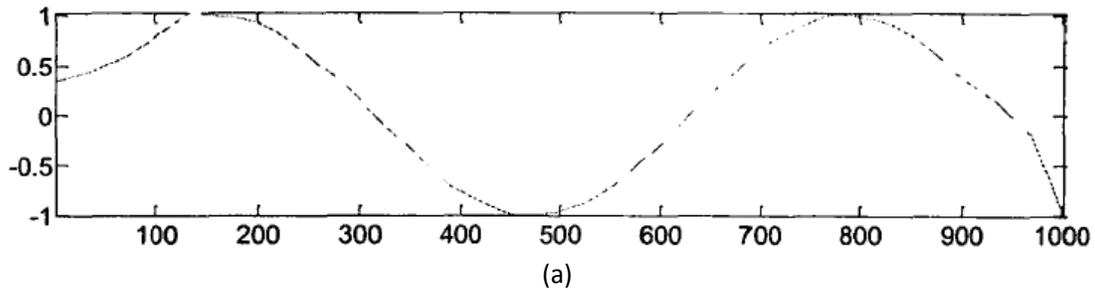
Table 1 shows the multi-scale entropy and energy of detailed hefts reconstruction in scale 1. It can be seen that in the scale 1, due to the impact of pulse signal, in the first 10 seconds, the scale entropy has the different of 11.267, however, the energy density changes only 0.001. The tiny changes could be easily ignored. But, the changes may cause the significant changes of the machine status even the machine is in his peak.

The other experiment is carried out by using another signal to present the machine status. The signal is  $y = \sin(t) + \sin(10t)$ , whose wave is shown in figure 3. It can be

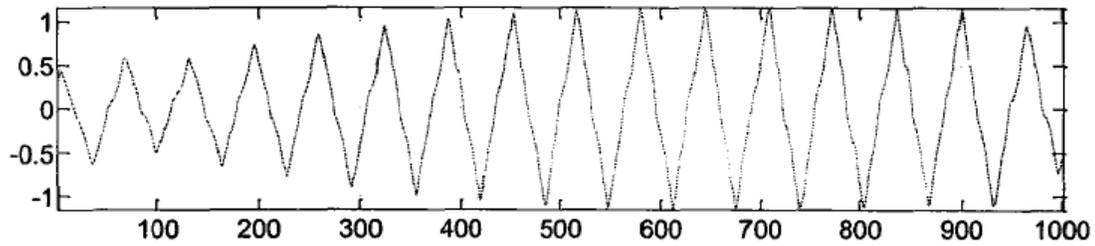
seen that the machine status is not as the previous experiment. The changes are significant within 10 seconds.



**Figure 3. Signal Wave**



(a)



(b)

**Figure 4. Multi-scale Analysis**

Figure 4 shows the multi-scale analysis of the above signal with the scale 6. Figure 4 (a) shows the similar hefts reconstruction, which shows the trends of the machine status changes. It is observed that, as the time increasing, the machine status will be unstable. The surveillance should be carried out every 500s which indicates the most probably machine breakdown will be occurred. Figure 4 (b) shows the detailed hefts reconstruction, which presents the changes of machine status. It is founded that, as the time increasing, the changes are becoming larger. At the beginning, machine status is stable. After 500s, the machine is becoming unstable. That could predict the machine status that should be a repair or maintenance so that the machine could work continuously.

**Table 2. Multi-scale Energy of Reconstructed Signal in Scale 6**

	Scale Energy
Detailed Hefts Reconstruction	548.234
Similar Hefts Reconstruction	345.210

Table 2 reports on the multi-scale energy of reconstructed signal in scale 6. The detailed hefts reconstruction is much bigger than the similar hefts reconstruction with the difference is 203.024. That indicates the trends from the scale are able to reflect the changes of the machine status. So that, this model is able to predict the machine changes and give some guidance for end users when carrying out machine maintenance or repair work.

## 6. Conclusions

This paper presents a method to carry out the machine surveillance from large number of data. Given the characteristics of big data in industrial field, this method proposes a multi-scale system with the entropy and energy to reflect the machine status in large scales. The multi-scale status surveillance models are introduced by using the residual analysis. This paper carries two experiments to test the method proposed and the experimental results show the feasibility and practicality of using the method in real-life case.

Future research work will be carried out in two aspects. Firstly, the noise of the signal or data could be removed so as to improve the accuracy of the method. Thus, a removing algorithm will be studied to add into the method before analyzing the data for decision-making. Secondly, the surveillance of machine status may involve in large number of factors, from which is the most important one or ones should be investigated. Therefore, this model could be improved that adding the influence weight of different factors so that different situations could be examined.

## References

- [1] B. Hopkins and B. Evelson, "Expand your digital horizon with Big Data, Forrester", vol. 30, (2011), September.
- [2] Y. Wang, Z. He, J. Xiang, and Y. Zi, "Application of local mean decomposition to the surveillance and diagnostics of low-speed helical gearbox", Mechanism and machine theory, vol. 47, (2012), pp. 62-73.
- [3] M. Spadacini, S. Savazzi, M. Nicoli, and S. Nicoli, Wireless networks for smart surveillance: Technologies, protocol design and experiments, in Wireless Communications and Networking Conference Workshops (WCNCW), IEEE, 2012, (2012), pp. 214-219.
- [4] R. Angeles, "RFID technologies: supply-chain applications and implementation issues", Information systems management, vol. 22, (2005), pp. 51-65.
- [5] R. Y. Zhong, Q. Y. Dai, K. Zhou, and X. B. Dai, "Design and Implementation of DMES Based on RFID," in 2nd International Conference on Anti-counterfeiting, Security and Identification, Guiyang, 20-23 Aug. 475-477, (2008), pp. 475-477.
- [6] R. Y. Zhong, G. Q. Huang, and Q. Y. Dai, "A Big Data Cleansing Approach for n-dimensional RFID-Cuboids", Proceeding of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2014), Taiwan, (2014) May 21-23.
- [7] R. Y. Zhong, G. Q. Huang, Q. Y. Dai, and T. Zhang, "Mining SOTs and Dispatching Rules from RFID-enabled Real-time Shopfloor Production Data", Journal of Intelligent Manufacturing. vol. 25, (2014), pp. 825-843.
- [8] R. Y. Zhong, G. Q. Huang, Q. Y. Dai, K. Zhou, T. Qu, and G. J. Hu, "RFID-enabled real-time manufacturing execution system for discrete manufacturing: Software design and implementation", in Networking, Sensing and Control (ICNSC), 2011 IEEE International Conference on, (2011), pp. 311-316.
- [9] R. Y. Zhong, Z. Li, A. L. Y. Pang, Y. Pan, T. Qu, and G. Q. Huang, "RFID-enabled Real-time Advanced Planning and Scheduling Shell for Production Decision-making", International Journal of Computer Integrated Manufacturing. , vol. 26, (2013), pp. 649-662.
- [10] Q. Z. Sheng, S. Zeadally, Z. Luo, J. Y. Chung, and Z. Maamar, "Ubiquitous RFID: Where are we?", Information systems frontiers, vol. 12, (2010), pp. 485-490.

- [11] P. Chaurand, D. S. Cornett, P. M. Angel, and R. M. Caprioli, "From whole-body sections down to cellular level, multiscale imaging of phospholipids by MALDI mass spectrometry", *Molecular & Cellular Proteomics*, vol. 10, (2011), p. O110. 004259.
- [12] V. Khanagha, K. Daoudi, O. Pont, and H. Yahia, "Improving text-independent phonetic segmentation based on the microcanonical multiscale formalism", in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, (2011), pp. 4484-4487.
- [13] Q. He and J. Wang, "Effects of multiscale noise tuning on stochastic resonance for weak signal detection," *Digital Signal Processing*, vol. 22, (2012), pp. 614-621.

## Author



**Fuzhen Xie**, she is a lecturer in School of Mechanical Engineering Xinyu University, Jiangxi Xinyu. She has published over 20 papers in international journals and conferences. Her research interests include automation control and application in industrial engineering.