Fault Diagnosis Based On Wavelet Fuzzy Feature Extraction and Information Fusion

MohammadReza Vazifeh , Farzaneh Abbasi Hossein Abadi

Ph.D. student in computer science and technology, Wuhan University of Technology-China,
Ph.D. student in Information engineering, Wuhan University of Technology-China

Abstract

With increasing demand for efficiency and product quality and progressing integration of automatic control systems in high cost and safety –critical process the field of supervision or monitoring, fault detection and diagnosis plays important rules. The Fault diagnosis task consist of the determination of the fault type with as many details as possible such as the fault size, location and time of detection. Today, fault diagnosis is main research in world. We exposure new algorithm in this paper, this algorithm have 3 steps. In the First step used wavelet packet and fuzzy set for make wavelet tree with coefficient in each node. In second step using wavelet tree and fuzzy set fused data with maximum Entropy coefficient in wavelet tree and in step three with output of fusion function we classification this fusion data. This algorithm have best time study because the time of search algorithms is 2^D ,D is depth of wavelet tree. Our proposed fusion strategies take into account that a Wavelet-Fuzzy by finding the optimal hyper plane with maximal margin. Then a Support Vector Machine classifier is trained. In the distributed schemes, the individual data sources are processed separately and modeled by using the Support Vector Machine.

Key Words:

Fault diagnosis, Information Fusion, Wavelet-Packed Entropy, Fuzzy set

I. Introduction

It is important to reduce maintenance costs and prevent unscheduled downtimes for machinery. So knowledge of what, where and how faults occur is very important. Condition-based maintenance (CBM) has the potential to decrease life-cycle maintenance costs, increase operational readiness and improve safety [1]. Fault detection and failure mode diagnosis are also necessary for implementing CBM (Byington and Garga, 2001).

Wavelet Transform (WT) has been developed. WT is a kind of variable window technology, which uses a time interval to analyze the high frequency and the low frequency components of the signal .The data using WT can be decomposed into approximation and detail coefficients in a multistate, presenting then a more effective tool for non-stationary signal analysis than the FT. Many studies present the applications of WT to decompose signals for improving the performance of fault detection and diagnosis in rotating machinery [2], [3], [4].

In this work, we propose to implement the WT for condition monitoring of rotating machinery. It is evaluated using the experimental measurements data in the cases of mass unbalance and gear fault. The main goal of this technique is to obtain more detailed information contained in the measured data [5], [6].

Support Vector Machines (SVMs), derived from statistical learning theory and VC-dimension theory, have been widely used in many fields and show good performance [7], [8].

SVMs were originally developed to solve binary classification problems, and so cannot be easily applied to diagnose more than one category of faults. In real world problems, discrimination between more than two categories is often required.

In an information processing system, fusion can take place at three levels: signal level, feature level, and decision level. Signal-level fusion is often used to reduce measurement uncertainty of a single sensor. Feature-level fusion can effectively use complementary information from different sources. One of the practical limitations is the tremendous size of the feature space and the resulting heavy computational burden. Therefore, feature−level and decision−level fusion are mainly considered in this paper [9], [10].

Fault diagnosis consists of 3-sub set:

- ✓ Feature Extraction
- ✓ Feature selection(Data Reduction)
- ✓ Fault Isolation (Classification), (fig 1).

II. Novelty algorithm for fault diagnosis

2.1 Feature extraction based Wavelet- fuzzy Entropy

Wavelet transform is a powerful technique in analyzing non-stationary signals such as PCG signals. The main advantage of wavelet transform is its varying window size that is narrow for high frequencies and wide for low frequencies. Therefore, wavelet transform is much more powerful than the other time frequency analysis techniques such as DFT and STFT, not only for providing

Manuscript received October 5, 2014
Manuscript revised October 20, 2014

useful time and frequency information, but also for its adaptive time and frequency resolution [3].

One of the quantitative measures associated with WPT is entropy. Entropy provides valuable information for analyzing non-stationary signals. The background theory of WPT and the definition of the entropy used in the current study are explained in this section [6].

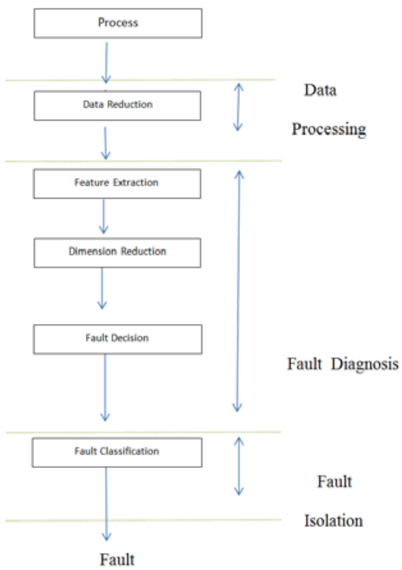

Fig 1: Fault Diagnosis framework

WPT is an extension of DWT whereby all nodes in the tree structure are allowed to split further at each level of decomposition. With WPT, both the approximation and detail coefficients are decomposed into approximation and detail components, in comparison to DWT that decomposes only the approximation coefficients of the signal. Therefore, features can be generated based on approximation and detail coefficients at different levels to obtain more information [7]. The WPT of a signal x (t) is defined as follows:

$$x_p^{n,j} = 2^{\frac{j}{2}} \int x(t)\varphi_n(2^{-j}t - p)dt,\ \ 0 \le n \le 2^s - 1 \quad (2\text{-}1)$$

Where n is the channel number, j is the number of decomposition level, or scale parameter, p is the position parameter, φ(t) is the mother wavelet, S is the maximum decomposition level [8]. After decomposing signal x (t) by WPT, 2^ssequences can be produced in the Sth level. The fast decomposition equation for this kind of WPT is:

$$x_k^{2n,j+2} = \sum_{p\epsilon z} h(p-2k)x_p^{n,j} \ . \quad (2\text{-}2)$$

$$x_k^{2n+1,j+1} = \sum_{p\epsilon z} g(p-2k)x_p^{n,j} \ . \quad (2\text{-}3)$$

In fuzzy sets, we allow any pattern $x_k \in R^n$ to depend to several classes to varying degrees. In this, the resulting class allocation becomes less strict than the one encountered in the crisp case .There are two classical theories of uncertainty within which we can define the entropy, the first and the most well-known is based on the notion of probability, while the second is based on the notion of possibility [12]. In the probabilistic approach, Shannon entropy is a well-known measure of uncertainty and is extensively covered in the literature. An extension to Shannon entropy is the concept of fuzzy entropy, in which fuzzy sets are used to aid the estimation of the entropies. It should be highlighted that the measuring of the fuzzy entropy is quite different from the classical Shannon entropy since fuzzy entropy contains fuzziness uncertainties (possibility), while Shannon entropy contains randomness uncertainties (probabilistic). However, a well-defined fuzzy entropy must satisfy the four Luca–Termini axioms [12] given as follows. The first task was to estimate the required memberships of the samples along each dimension (or feature) in all of the problem classes. Several methods were reported in the literature for estimating the membership functions including the kNN approach [13], the well-known fuzzy c-means method (employed within the FWP) [14], and many other variants [15], [16]. Given the high computation cost associated with the KNN near and the singularity problem of the fuzzy c-means, we offer to use the following approach for estimating the membership values.

Given a universal set with elements x_k distributed in a pattern space as $X = \{x_1, x_2, \ldots, x_l\}$, where k = 1, 2, . . . , l with l being the total number of patterns. For easy, it will be useful to describe the membership value that the k th vector has in the ith class with the following record:

$$\mu_{ik} = \mu_i(x_k) \in [0,1] \quad (2\text{-}4)$$

Mark the mean of the data sample that belong to class i as x_i and the radius of the data as r

$$r = \max||\bar{x}_i - x_k||_\sigma \quad (2\text{-}5)$$

Then the fuzzy membership μ_{ik} can be calculated:

$$\mu_{ik} = \left(\frac{||\bar{x}_i - x_k||_\sigma}{r+\epsilon}\right)^{\frac{-2}{m-1}} \quad (2\text{-}6)$$

Where m is the fuzzian parameter, and $\epsilon > 0$ is a small value to avoid singularity, and σ is the standard deviation involved in the space computation. Finally, the membership of each of the samples in all of the problem classes is normalized according to $\sum_{i=1}^{c} \mu_{ik} = 1$

Let $X = \{x_1, \ldots, x_n\}$ be a discrete random variable with a finite set consist of n symbols, and let $\mu_A(x_i)$ be the fellowship degree of the element xi to fuzzy set A, and F be a set-to-point mapping $F: G(2^x) \to [0,1]$. Hence, F is a fuzzy set determined on fuzzy sets. F is an entropy measure if it satisfies the following Luca–Termini axioms [17], [18]:

$F(A) = 0$ Iff $A \in 2^x F(A) = 0$

Where A is a non-fuzzy set and 2^x indicates the power set of set A.

$F(A) = 1$ Iff $\mu_A(x_i) = 0.5$ for all i;

$F(A) \le F(B)$ if A is less fuzzy than B (for example if $\mu_A(x_i) < \mu_B(x_i)$ when $\mu_B(x_i) \le 0.5$ and $\mu_A(x_i) \ge \mu_B(x_i)$ when $\mu_B(x_i) \ge 0.5$;

$$F(A) = F(A^c)$$

$$A^c = (1 - \mu_A(x_1), \dots, 1 - \mu_A(x_n))$$

Shannon entropy satisfies the aforesaid four De Luca–Termini axioms, where for a discrete random variable X with a probability mass function $p(x_i)$, Shannon entropy is defined by:

$$H(x) = -\sum p(x_i) \log_2^{p(x_i)} \quad (2\text{-}7)$$

Using the proposed membership function in (12), we create c-fuzzy sets along each specific feature f , each feature will be in turn reflect the membership degrees of the samples in each of the c problem classes. The fuzzy equivalent to the joint probability of the training patterns that belong to class i is given here as

$$P(f, c_i) = \frac{\sum_{k \in A_i} \mu_{ik}}{\text{Total Pattern}} \quad (2\text{-}8)$$

Where $P(f, c_i)$ can be translate as the degree by which the samples that are predefined to belong to class i does actually contribute to that specific class. A_i is the set of indexes of the training patterns belonging to class i.

The fuzzy entropy of the elements of class i, denoted as $H(f, c_i)$, is then equal to

$$H(f, c_i) = -\sum p_{f,c_i} \log_2^{p_{f,c_i}} \quad (2\text{-}9)$$

In order to account for the entropy along all c-classes, the common aforesaid entropy has to be summed along the universal set to generate the complete fuzzy entropy H (f ,C)

$$H(f, C) = \sum_{i=1}^{c} H(f, C_i) \quad (2\text{-}10)$$

The aforesaid entropy satisfies the four De Luca– Termini axioms and is termed as the joint fuzzy entropy. The aforesaid equations can be applied on the samples along each feature, thus computing the entropies associated with each feature.

In order to find the marginal entropy H (f) of each feature, we add the compute membership values of the samples along each of the c-fuzzy sets Si as follows:

$$P(fs_i) = \frac{\sum_{k \in A_i} \mu_{ik}}{\text{Total Pattern}} \quad (2\text{-}11)$$

Then the marginal entropy is found by:

$$H(f) = -\sum p_{fc_i} \log_2^{p_{fc_i}} \quad (2\text{-}12)$$

2.2 Information Fusion Techniques

The information fusion based on wavelet-fuzzy is to make wavelet packet transformation (WPT) to preparation data fusion and decompose it into different resolution space. Activity measure can acquire certain feature information of the multi-sensor analysis coefficient of the input, and decide which input has more obvious feature information. The general activity measure is a determine function relative to detail component amplitude. The definition is :

$$a_j^{\varepsilon}(m, n) = \sum_x p(m + m', n + n') * |D_j^{\varepsilon}(m + m', n + n')|^k \quad (2\text{-}13)$$

D_j^{ε} is the detail component coefficient matrix and $u_j^{\varepsilon}(m, n)$ is the activity measure of $D_j^{\varepsilon}(m, n)$, p is the mask of window area and it is used to linear filter D_j^{ε} ,

The activity dimensions said above are calculated by the components of detailed components decomposed with every level fail the impact of its corresponding approximate components. So the suggestion of fuzzy activity measure, taking the impact of both detail and approximate components on the activity measure into attention, achieves the objective of improving the effect of fusion.

Suppose p is the window mask of j th level's detail component:

$$p_{i,j} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,m} \\ \vdots & \ddots & \vdots \\ p_{n,1} & \cdots & p_{n,m} \end{bmatrix} \quad (2\text{-}14)$$

For every $p_{i,j}$.

Suppose in j level approximation coefficient matrix $p_{i,j}$ is

$$v_{i,j}^{l} = \begin{bmatrix} v_{1,1} & \cdots & v_{1,m} \\ \vdots & \ddots & \vdots \\ v_{n,1} & \cdots & v_{n,m} \end{bmatrix} \quad (2\text{-}15)$$

Then for the j th level detail component of input, the basic decision making module adopted with information fusion algorithm is:

$$\omega_{A,j}^{\varepsilon}(m,n) = \begin{cases} 0 & M_{j,AB}^{\varepsilon}(m,n) \le \sigma \text{ and } u_{j,A}^{\varepsilon}(m,n) \le u_{j,B}^{\varepsilon}(m,n) \\ 1 & M_{j,AB}^{\varepsilon}(m,n) \le \sigma \text{ and } u_{j,A}^{\varepsilon}(m,n) > u_{j,B}^{\varepsilon}(m,n) \\ \frac{1}{n} - \frac{1-M_{j,AB}^{\varepsilon}(m,n)}{1-\sigma} & M_{j,AB}^{\varepsilon}(m,n) \ge \sigma \text{ and } u_{j,A}^{\varepsilon}(m,n) \le u_{j,B}^{\varepsilon}(m,n) \\ \frac{1}{n} + \frac{1-M_{j,AB}^{\varepsilon}(m,n)}{1-\sigma} & M_{j,AB}^{\varepsilon}(m,n) \ge \sigma \text{ and } u_{j,A}^{\varepsilon}(m,n) \ge u_{j,B}^{\varepsilon}(m,n) \end{cases} \quad (2\text{-}16)$$

The superscript ε is the directions that detail component represents, $\omega_{A,j}^{\varepsilon}(m, n)$ is the decision factor of the fusion algorithm, $u_{j,s}^{\varepsilon}(m, n)$ is the fuzzy activity measure as described and σ is relativity threshold value. $M_{j,AB}^{\varepsilon}(m, n)$ is the relative co efficiency of the input

$$M_{j,AB}^{\varepsilon}(m, n) = \frac{\sum\sum u_{j,A}^{\varepsilon}(m+m', n+n') u_{j,B}^{\varepsilon}(m+m', n+n')}{\sum\sum |u_{j,A}^{\varepsilon}(m+m', n+n')|^2 + |u_{j,B}^{\varepsilon}(m+m', n+n')|'} \quad (2\text{-}17)$$

The wavelet fuzzy coefficient after fusion could be showed as:

$$D_j^{\varepsilon}(m, n) = \sum_{i=1}^{i \ne j} \omega_{j,i}^{\varepsilon}(m, n) U_{j,i}^{\varepsilon}(m, n) \quad (2\text{-}18)$$

In formula: $\sum_{i=1}^{n} \omega_{j,i}^{\varepsilon}(m, n) = 1$ and $U_j^{\varepsilon}(m, n)$ is detail component coefficient on j level, in the direction of ε .

2.3 SVM model

The classification problem can be restricted to consideration of the two-class problem without loss of generality. The goal is to produce a classifier that will work well on unseen examples. Here there are many possible linear classifiers that can separate the data, but there is only one that maximizes the margin (maximizes the distance between it and the nearest data point of each class). This linear classifier is termed the optimal separating hyper plane. Intuitively, we would expect this boundary to generalize well as opposed to the other possible boundaries [7].Consider the problem of separating the set of training vectors belonging to two separate classes,

$D = \{(x^i, y^i): x^i \in R^n, y^i \in \{-1,1\}\}$ (2-19)

With a hyper plan $< w, x > +b = 0$ (2-14)

The set of vectors is said to be optimally separated by the hyper plane if it is separated without error and the distance between the closest vectors to the hyper plane is maximal [1].There is some redundancy in Equation 214, and without loss of generality it is appropriate to consider a canonical hyper plane where the parameters w, b is constrained by:

$\min\left|< w, x^i > +b\right| = 1$ (2-20)

A separating hyper plane in canonical form must satisfy the following constraints,

$y^i[< w, x^i > +b] \geq 1 \quad i = 1, \dots, l$ (2-21)

The distance d (w, b; x) of a point x from the hyper plane (w, b) is $d(w, b; x) = \frac{|<w,x^i>+b|}{||w||}$ (2-22)

$\emptyset(w, b, a) = \frac{1}{2}||w||^2 - \sum_{i=1}^{l} a_i(y^i[< w, x^i > +b] - 1)$ (2-23)

Where α is the Lagrange multipliers, The Lagrangian have to be minimized with respect to w, b and maximized with respect to $\alpha \geq 0$. Classical Lagrangian duality enables the primal problem, Equation 2.17, to be transformed to its dual problem, which is easier to solve. The dual problem is given by,

$\max\ w(\alpha) = \max(\min \emptyset(w, b, a))$ (2-24)

The minimum with respect to w and b of the Lagrangian,∅ is given by,

$\frac{d\emptyset}{dt} = 0 \rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0$ (2-25)

$\frac{d\emptyset}{dw} = 0 \rightarrow w = \sum_{i=1}^{l} \alpha_i y_i x_i$ (2-26)

Hence from Equations 2.17, 2.18 and 2.21, the dual problem is : $\max w(\alpha) = \max \frac{-1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j < x_i, x_j > + \sum_{k=1}^{l} \alpha_k$ (2-27)

Solving Equation 2.21 with constraints Equation 2.17 determines the Lagrange multipliers, and the optimal separating hyper plane is given by

$$w^* = \sum_{i=1}^{l} \alpha_i y_i x_i \qquad (2-28)$$

$$b^* = -\frac{1}{2} < w^*, x_r + x_s >$$

Where x_r and x_s are any support vector from each class satisfying

$$\alpha_r, \alpha_s > 0, y_r = -1, y_s = 1 \qquad (2-29)$$

The hard classifier is then,

$$f(x) = \text{sgn}\,(< w^*, x > +b) \qquad (2-30)$$

A soft classifier may be used which linearly interpolates the margin [1, 3]

$f(x) = h(< w^*, x > +b)$ where $h(z) =$

$$\begin{cases} -1 & : z < -1 \\ 0 & : -1 \leq z \leq 1 \\ 1 & : z > 1 \end{cases} \qquad (2\text{-}31)$$

III. New algorithm

In new method we have 3 steps:

Step1: feature extraction

In this step have 3 subsets:

1-1: Wavelet Packet Transform

1-2: fuzzy set

In this section will be make wavelet-Tree with coefficient using fuzzy set.

Step2: Information Fusion based Wavelet Packet coefficient

Step3: classification using SVM with max wavelet tree.

The new method of identifying current from internal fault is obtained from Multi feature and wavelet packet transform using fuzzy set principle and classification using SVM method (Figure 2).

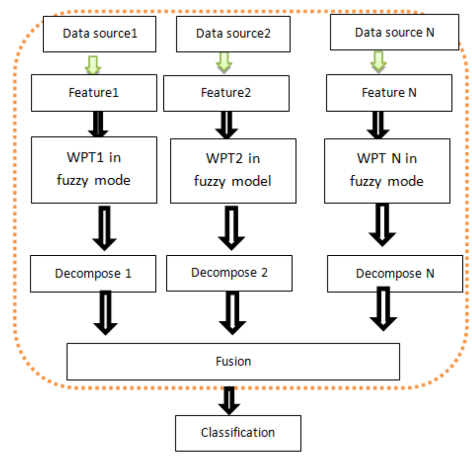

Fig 2: New algorithm schema

In this method we have N input (multi feature extraction) and using wavelet-entropy and information fusion technology we reduction dimension of input then using SVM we used for classification. In this hybrid system and new method have good result for fault diagnosis and we show in example that these have best method.

IV. Test algorithm with data

In this section we want test new method, it have 4 phases include Input-Learn-Train -Test.
We get data for my research idea from teacher huang in mechanic engineering school in Wuhan University of technology. This data get from 2 sensors in gearbox simulation in lab, then convert data to data file enable in signal processing programming environment (matlab).

a. Wavelet-Entropy ->wavelet-tree with coefficient node

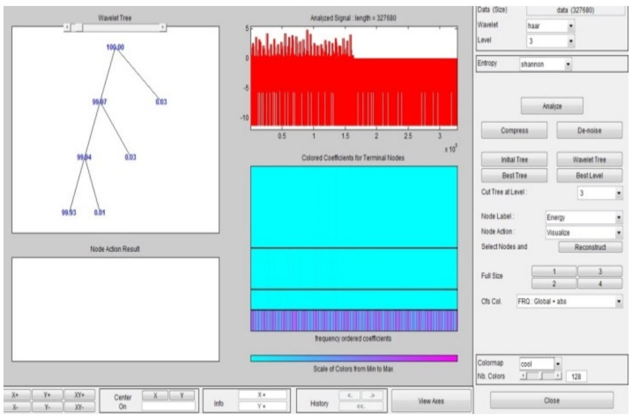

Figure 3: wavelet tree using WPD and Shannon Entropy

b. Fusion data

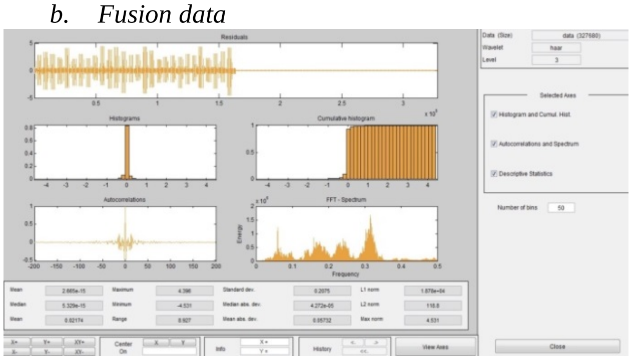

Figure 4: residual wavelet packet

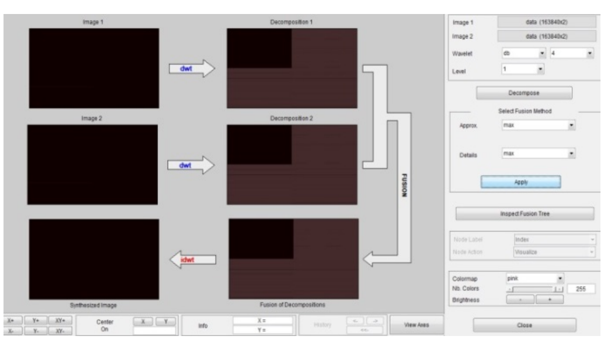

Figure 5: Data fusion wavelet packet

c. SVM Train:

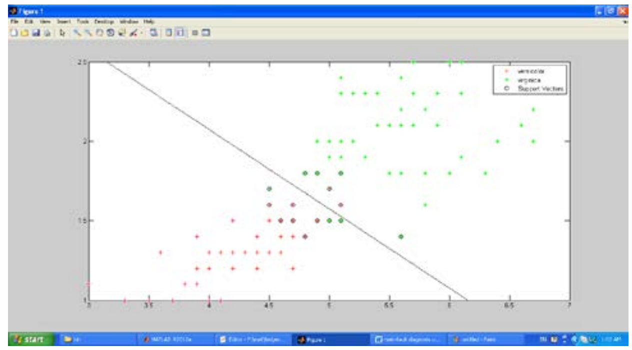

Fig 6: SVM classification

d. Test

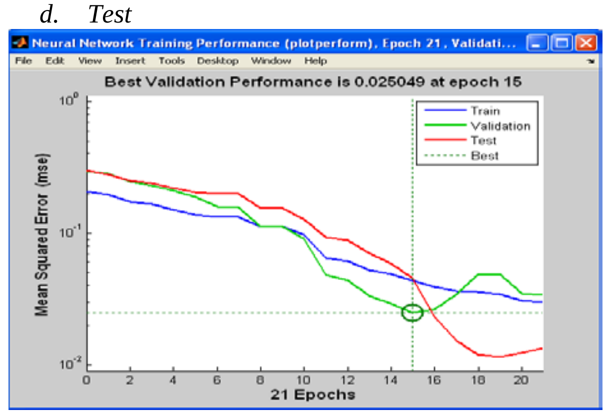

Fig 7: Test, Validation,Train (svm)

Conclusion

We described new algorithm in this paper, this new algorithm have 3 steps. In the First step used wavelet packet for make wavelet tree with coefficient in each node. In second step using wavelet tree and fuzzy set fused data with maximum coefficient in wavelet and in step three with output of fusion function we classification this fusion data.
In this algorithm we have best time study because the time of search algorithms is 2^D , D is depth of wavelet tree(decision tree) and fusion technique is linear model(maximum Entropy selection), SVM classification is supervised learning (linear classification).

References

[1] K. Shibata, A. Takahashi, T. Shirai, "Fault diagnosis of rotating machinery through visualization of sound signal," Journal of Mechanical Systems and Signal Processing, vol. 14, pp. 229-241, 2000.
[2] J. D. Wu, C.-H. Liu, "Investigation of engine fault diagnosis using discrete wavelet transform and neural network," Expert Systems with Applications, vol. 35, pp. 1200-1213, 2008.
[3] S. H. Cao, J. C. Cao, "Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis," Applied Thermal Engineering, vol. 25, no. 2-3, pp. 161-172, 2005.
[4] G. Tlusty, Manufacturing Processes and Equipment. Upper Saddle River, NJ: Prentice-Hall, 2000.
[5] V.N. Vapnik, Statistical Learning Theory, New York: John Wiley & Sons, 1998.
[6] J. E. Seem and K. D. Knissmann, "Statically methods for on-line fault detection in press-working applications," presented at the ASME Winter Annual Meeting, 1994.
[7] W.S. Chen, P.C. Yuen, J. Huang, and J. Lai, "Face classification based on shannon wavelet kernel and modified fisher criterion," in Proc. 7th International Conference on Automatic Face and Gesture Recognition (FGR 2006), April 2006, pp. 467–474.
[8] L. Zhang, W. Dazhou, and L.C. Jiao, "Wavelet support vector machine," IEEE Tran. Systems, Man, and Cybernetics, vol. 34, no. 1, pp. 34–39, 2004.
[9] S.G. Nikolov, D.R. Bull, C.N. Canagarajah, M. Halliwell, P.N.T. Wells, Fusion of 2-D images using their multiscale edges, Proceedings of the International Conference on Pattern Recognition, pp. 1051–1054,2000.
[10] Rockinger, Image sequence fusion using a shift-invariant wavelet transform, Proceedings of the IEEE International Conference on Image Processing, Vol. III, pp. 288–291,1997.
[11] Greenspan H, Ruf A, Goldberger J. Constrained Gaussian mixture model framework for automatic segmentation of MR brain images. IEEE Trans. Med. Image. 25: 1233-45,2006.

MOHMMADREZA VAZIFEH,PAN HAO,Fault Diagnosis based wavelet-entropy and Information Fusion, International Conference on Computer Science and Automation Engineering, Guangzhou, China, in process