Journal of Intelligent Manufacturing 12, 257–268, 2001

Automatic feature extraction of waveform signals for in-process diagnostic performance improvement

JIONGHUA JIN[1] and JIANJUN SHI[2]

[1]Department of Systems and Industrial Engineering, The University of Arizona, P.O. Box 210020, Tucson, Arizona 85721-0020
[2]Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109-2117

Received August 1999 and accepted April 2000

In this paper, a new methodology is presented for developing a diagnostic system using waveform signals with limited or with no prior fault information. The key issues studied in this paper are automatic fault detection, optimal feature extraction, optimal feature subset selection, and diagnostic performance assessment. By using this methodology, a diagnostic system can be developed and its performance is continuously improved as the knowledge of process faults is automatically accumulated during production. As a real example, the tonnage signal analysis for stamping process monitoring is provided to demonstrate the implementation of this methodology.

Keywords: Automatic feature extraction, Haar transform, waveform signals, process monitoring, fault diagnosis

1. Introduction

Monitoring and diagnostic systems have played an important role in modern manufacturing process control. Many intelligent or knowledge-based systems have been successfully developed for different application domains. However, the development of such a system normally requires sufficient prior knowledge or fault condition data, which is hard to satisfy in manufacturing systems. This is especially true for a new product or process launch. Thus, the motivation of the research presented in this paper is to address this important issue by developing a methodology for monitoring and diagnostic system development with limited or with no prior fault information. Waveform signals are used as the essential information for the diagnostic system development.

Waveform signals represent a class of analog or digital signals over time, which normally can be measured using in-process sensors in a manufacturing process. It has broad potential applications, such as tonnage signals in stamping, torque signals in tapping, and force signals in welding, which are shown in Fig. 1 (a)–(c), respectively. In general, those waveform signals contain rich information that can be related to both product quality and process variables. The characteristics of those waveform signals studied in this paper are summarized as follows:

- Non-stationary.
- Working cycle-based signals, meaning that each cycle of a waveform signal covers a complete cycle of an operation.
- Segmental signals, meaning that different segments represent different process stages, which may have different potential process faults.
- Localized time and frequency components in different segments.
- In-process automatic sensing.

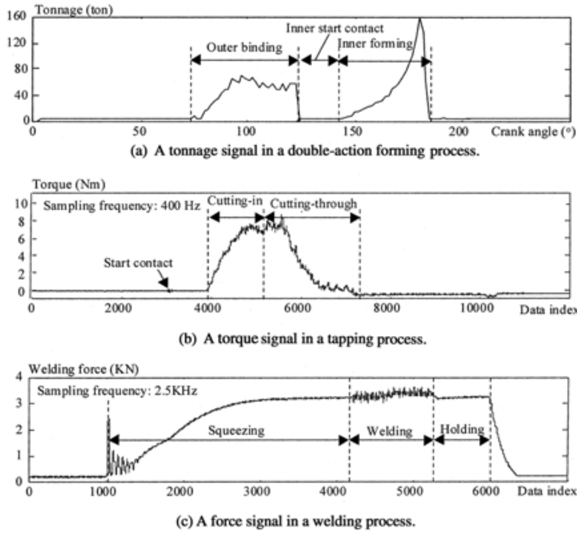

Fig. 1. Waveform signals measured in the different production processes.

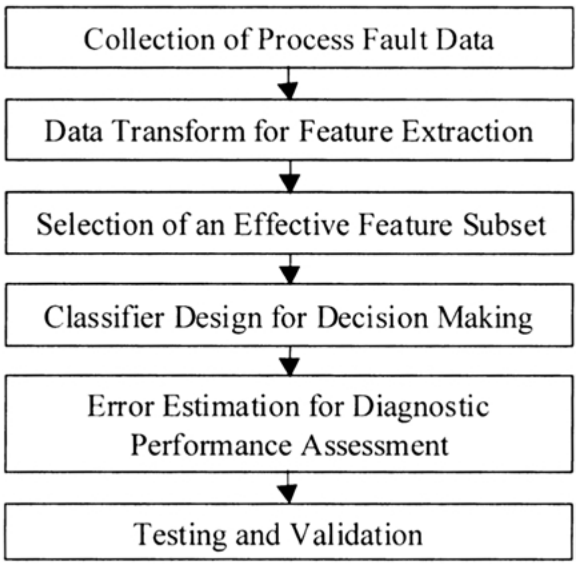

Fig. 2. Procedures for a feature-based fault diagnosis system development.

There are two basic approaches in diagnostic system development by using waveform signals: a model-based approach and a feature-based approach. In the model-based approach, observations are considered as a time-ordered stochastic process. The critical concern of using this approach is to have an appropriate process model which is sensitive to process faults but robust to process noises (Deibert and Holfling, 1992). In addition, fault models or fault signal characteristics need to be known before making fault detection. However, these types of information are normally not available at the beginning of the production due to the complex relationship between waveform signals and the associated manufacturing process. Thus, a model-based approach is generally not effective when waveform signals are used for diagnostic system development.

A feature-based approach is more suitable to a complex process where waveform signals are used for process diagnosis. In such a system, features are considered as random variables or as a random set. Feature extraction and feature subset selection are critical steps to reduce the number of attributes or data dimension considered in the decision-making step (Kharin, 1992). The conventional procedure to develop a feature-based diagnostic system is shown in Fig. 2. The essential requirement, or precondition, to use this conventional procedure is that sufficient historical fault data or prior fault knowledge are available before developing a diagnostic system. In many applications, this precondition is not satisfied especially during the new machine or process launch. Therefore, it is a challenging problem to develop a feature-based diagnostic system with limited or with no prior fault information.

In this paper, we will propose an automatic feature extraction methodology for the development of a feature-based diagnostic system using waveform signals. In this methodology, the wavelet analysis is used as a basic tool for feature extraction. The wavelet analysis is selected for this research due to its multiresolution nature, its localized properties in both time and frequency domains, its fast algorithms ready for an on-line implementation, and its efficient data compression for feature extraction. The Haar transform is selected in the paper because it has an explicit geometrical interpretation for a detected change of a Haar coefficient. These interpretations can be easily associated with the profile change of a waveform signal due to process faults.

Wavelet analysis has been widely used in image and speech processing for decades. Much research has been focused on the data shrinkage and signal noise filtering (Coifman and Yale, 1992; Donoho and Johnstone, 1994). The wavelet transform used as a feature extraction method has recently received more and more attention for process monitoring in manufacturing processes, such as drill condition monitoring (Tansel *et al*., 1993), face milling failure detection (Kasashima *et al.*, 1995), and spalling detection on ball bearings (Mori *et al.*, 1996). In these applications, the feature extraction method or wavelet coefficient (or a function of them) selection is mostly based on engineering knowledge, or based on the use of conventional trial-and-error approaches when sufficient prior fault data are available. The relevant coefficients associated with characteristic frequencies of a dynamic system are usually selected as features. However, for a complex manufacturing process, there is no sufficient prior knowledge to describe the complex relationship between waveform signals and process faults. Thus, it is very difficult to pre-determine the process characteristic frequencies, and the trial-and-error approach cannot be implemented.

This paper presents a new methodology for developing a diagnostic system with limited or with no prior fault information. In this methodology, a monitoring decision for detecting process faults is made first based on normal production condition. Then, the detected fault is further classified, and the knowledge of process faults is continuously accumulated during the use of this diagnostic system in production. When a new process fault is found, the optimal feature subset is adaptively updated to include the new characteristics of the newly detected fault. Thus, the diagnostic capability could be continuously improved as new fault data are automatically accumulated and classified.

As can be seen, the proposed methodology is to emphasize how to continuously improve diagnostic ability through machine learning. The major research issues discussed in the paper show the common research problems of applying artificial intelligence to machine learning for the process monitoring and fault diagnosis purpose. An adaptive supervisory learning method is used in this paper for feature extraction and fault classification. The capability of adaptive learning of process fault knowledge during routine manufacturing production can be seen from the following two aspects: (1) The increase of the fault samples in the known fault clusters can fine-tune the parameters of an existing classifier to improve its diagnostic accuracy; and (2) The addition of the newly identified fault patterns can enhance its diagnostic ability for more process fault diagosis. More discussions on these features will be given in Section 3.

The outline of this paper is listed as follows. An overview of the new methodology is given in Section 2. Then, Section 3 provides detailed research procedures for the methodology development. After that, a real example of stamping processes is provided in Section 4. Finally, a summary and information about future work are given in Section 5.

2. Methodology overview

The novel idea of this paper is to develop a methodology for developing a diagnostic system by using waveform signals with limited or with no prior fault information. The basic principles and generic procedures are shown in Fig. 3. In this methodology, there is no requirement of prior knowledge of process faults. Thus, in-process information assessment, fault classification, and fault knowledge accumulation are very critical. The essential idea of this method is that more and more knowledge about process faults can be accumulated during the use of the process monitoring system. That knowledge is then used to further improve the diagnostic performance. For this purpose, in-process information assessment needs to be addressed in step 1 to judge whether the current measurement reflects a normal working condition, a known fault, or a new fault. If the current measurement represents a normal working condition or a known fault condition, it will be added into the cluster

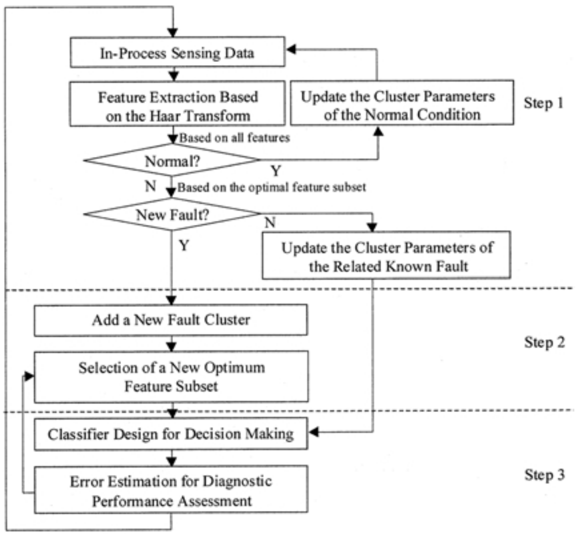

Fig. 3. The framework of diagnostic system development with adaptive learning.

of the normal working conditions or the known fault conditions respectively. The inclusion of this new sample will increase the sample size in the associated cluster, thus improving its parameter estimation accuracy. If it is a new fault, it will form a new fault cluster and get into step 2 of the development. In order to represent the knowledge of process faults effectively, automatic feature extraction and optimal feature subset selection will be implemented by maximizing the separability of all known fault clusters. Finally, classification errors obtained from decision making are analyzed in step 3 by using the cross-validation method based on the selected feature subset. This classification error will be used as a criterion to determine the optimal dimension of a feature subset in terms of the improved in-process diagnostic performance.

In order to develop this new methodology, the following issues will be discussed in the paper: (1) Wavelet Haar transform used for effective feature extraction in terms of signal characteristic representation (Section 3.1); (2) In-process sensing data assessment, which detects whether the process is normal, and if it is abnormal, whether it is a new fault or a known fault (Section 3.2); (3) Classifier design and decision-making for multiple fault classification (Section 3.3); (4) Optimal feature subset selection by maximizing fault cluster separability for diagnostic performance improvement (Section 3.4); (5) Classification error estimation via cross validation for diagnostic performance assessment (Section 3.5). The detailed procedures and algorithms will be discussed in the following sections.

3. Automatic feature extraction and classification

3.1. *Harr transform for feature extraction*

The first step in the methodology development is data transformation for feature extraction. A wavelet transform is chosen as an effective tool for the applications where a signal profile change is a major concern. When a signal is non-stationary and has localized frequency components, the wavelet multi-resolution analysis (MRA) can be used to decompose the signal into different resolutions. Moreover, the orthogonality and compactness of wavelets will ensure an efficient data compression to remove the noise and other process irrelevant signals from a waveform signal. As a result, the number of features, i.e., the number of compressed wavelet coefficients, is much less than that of the original data point.

The Haar transform is chosen as a data transform here because Haar coefficients have an explicit interpretation of the changes in a waveform profile. A brief review of the Haar transform is given as follows.

The Haar transform $h(r,q,t)$ is the first order Daubechies wavelet function (Daubechies, 1992). $h(0,0,t)$ is a Haar scale function defined in $[0,1]$ by

$$h(0,0,t) = 1, \quad t \in [0,1] \tag{1}$$

and $h(r,q,t)$ $(r,q \geq 1)$ is a Haar wavelet function which is obtained through the dilation and translation of the Haar mother wavelet function $h(1,1,t)$ by

$$h(r,q,t) = \begin{cases} 2^{\frac{r-1}{2}}, & \frac{q-1}{2^{r-1}} < t < \frac{q-1/2}{2^{r-1}} \\ -2^{\frac{r-1}{2}}, & \frac{q-1/2}{2^{r-1}} < t < \frac{q}{2^{r-1}} \\ 0, & \text{elsewhere}, \quad \forall t \in [0,\ 1] \end{cases} \tag{2}$$

where r represents the decomposition scale, q is the number of functions within scale r.

When a discrete Haar transform $\mathbf{H}$ is used for a discrete waveform signal $\mathbf{Y}$ ($\mathbf{Y}$ is a vector with N data point, i.e., $\mathbf{Y} \in \mathbf{R}^{N\times 1}$), the Haar coefficients $\mathbf{C}$ can be obtained by

$$\mathbf{C} = \mathbf{HY} \tag{3}$$

where $\mathbf{H} \in \mathbf{R}^{N\times N}$ is a matrix representing a discrete Haar transform function. $N = 2^n$ and n is a positive integer representing the maximum decomposition scale. Each row of $\mathbf{H}$ is a discrete Haar function obtained by sampling the continuous Haar function $h(r,q,t)$ as shown in Equation 2. There are 2^{r-1} rows for a given decomposition scale $r(1 \leq r \leq n)$. $h_{i,j}(1 \leq j \leq 2^n)$ is the ith row of H, which is generated by Strang and Nguyen (1996):

$$h_{i,j} = 2^{-n/2} h(r,q,j\cdot 2^{-n}) \tag{4}$$

where if $i=1$, then $r=0$ and $q=0$; if $2 \leq i \leq 2^n$, then $i = 2^{r-1}+q$ with $1 \leq r \leq n$ and $1 \leq q \leq 2^{r-1}$. The coefficient $2^{-n/2}$ is used to normalize each row of H. The explicit relationship between Haar coefficients $\mathbf{C} = [c_0^0, c_1^1, c_2^1, c_2^2, c_3^1, \ldots, c_n^{2^{n-1}}]^T$ and the mean values of the subset data over $\mathbf{Y} = [y_1, y_2, \ldots, y_{2^n}]^T$ can be seen as follows:

$$\text{For } r=0, \qquad c_0^0 = 2^{n/2}\bar{Y}(1,2^n) \tag{5}$$

For $r \geq 1$,

$$c_r^q = 2^{\frac{n-r-1}{2}}\left\{\bar{Y}\left[(q-1)2^{(n-r-1)}+1,\right.\right.$$
$$\left.\left(q-\frac{1}{2}\right)2^{(n-r+1)}\right]-\bar{Y}\left[\left(q-\frac{1}{2}\right)2^{(n-r+1)}+1,$$
$$\left.\left.q2^{(n-r+1)}\right]\right\} \quad (6)$$

where $\bar{Y}[i,j]$ is the mean within data region $[i,j]$, that is: $\bar{Y}[i,j]=1/(j-i+1)\sum_{k=i}^{j} y_k$.

The explicit geometrical interpretation of Haar coefficients is that c_r^q, except for c_0^0, is proportional to the mean difference between two adjacent intervals. Moreover, the higher the decomposition scale r, the higher the frequency component the Haar function represents. Therefore, Haar coefficients obtained at the different scales can represent the mean differences under the different decomposition resolutions. In the following study, the Haar coefficients c_r^q remaining after data denoising (Donoho and Johnstone, 1994) will be considered as the whole feature space to be used for process monitoring and fault classification.

3.2. *In-process sensing data assessment*

3.2.1. *Hotelling T^2 control limits for process normal condition monitoring*

The waveform signal studied in this paper is a working cycle-based signal. Each cycle corresponds to a complete manufacturing operation, which contains p-correlated random variables $\mathbf{x}_i=[x_{i1},x_{i2},\ldots,x_{ip}]^T$. x_{ij} represents the ith cycle measurement at time j $(j=1,\ldots,p)$. Under normal working conditions, the baseline of waveform signals is assumed to follow a p-dimensional multivariate normal distribution $N(\boldsymbol{\mu}(0),\Sigma(0))$, where $\Sigma(0)$ is a $p\times p$ covariance matrix of p random variables for class 0 (class 0 is referred as a normal working condition), and $\mu(0)$ is a mean vector of class 0. In order to simplify the notation, the class notation 0 is ignored in this subsection. Generally, the class parameters $\boldsymbol{\mu}$ and Σ and unknown and need to be estimated from normal working conditions (or in-control conditions).

Alt (1985) has pointed out that there are two distinct phases in constructing Hotelling T^2 control limits. The objective of phase 1 is to obtain a set of observations which are in-control, and then μ and Σ are estimated using these in-control data. For this purpose, a set of preliminary samples is analyzed to establish a phase-1 control limit. This control limit is then used to check whether the preliminary samples are in-control. After eliminating all out-of-control samples from the preliminary samples, the obtained $\hat{\mu}$ and $\hat{\Sigma}$ will be used to establish phase-2 control limits for process monitoring.

Phase 1: Establish a control limit based on preliminary data

Hotelling (1947) provided the first solution to the multivariate detection problem by suggesting the use of the statistic T^2. For a single observation, the statistic is defined as

$$T^2=(\mathbf{x}_i-\bar{\mathbf{x}})^T\mathbf{S}^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}) \quad (7)$$

where $\bar{\mathbf{x}}$ is the estimated mean vector,

$$\bar{\mathbf{x}}=[\bar{x}_1,\bar{x}_2,\ldots,\bar{x}_p]^T \quad (8)$$

and $\bar{x}_j=(1/m)\sum_{i=1}^{m}x_{ij}$ is the estimate of mean for variable j $(j=1,\ldots,p)$, and m is the number of subgroups. $\mathbf{S}$ is the estimated covariance matrix given by

$$\mathbf{S}=\frac{1}{m-1}\sum_{i=1}^{m}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^T \quad (9)$$

At phase 1, $\bar{\mathbf{x}}$ and $\mathbf{S}$ need to be estimated for phase 2 based on the preliminary in-control data. In order to obtain these in-control data from the preliminary data, the respective phase-1 control limit needs to be established. It has been proven that statistic T^2 follows an F distribution, which is provided by the Appendix in the paper. Thus, under a given Type 1 error α, the phase-1 upper control limit UCL_1 can be established by

$$\mathrm{UCL}_1=\frac{p(m-1)^2}{m(m-p)}F_\alpha(p,m-p) \quad (10)$$

where $F_\alpha(p,m-p)$ is the $1-\alpha$ percentile of the F distribution with p and $m-p$ degrees of freedom. In the paper, it is assumed that the change of statistic T^2 is mainly due to a process mean shift. So, the lower control limit is set to zero because any mean shift will lead to an increase in the statistic T^2.

Based on the established phase-1 control limit, the outliers due to assignable causes can be removed from the preliminary data set if they are out of the phase-1 upper control limit. Afterward, a re-estimation of the control limit is performed, and the remained data are rechecked by the revised control limit. This iterative checking procedure is continuously repeated until all remained preliminary data are in-control according to the phase-1 control limit.

Phase 2: Establish control limits for process normal condition monitoring

At phase 2, statistic T^2 of a waveform signal $\mathbf{x}_i$ is calculated based on $\bar{\mathbf{x}}$ and $\mathbf{S}$ estimated at phase 1. Although statistic T^2 still follows an F distribution, it is multiplied by a different constant because $\mathbf{x}_i$ at phase 2 is always independent of the estimates of $\bar{\mathbf{x}}$ and $\mathbf{S}$ (Tracy *et al.*, 1992). Thus, for a given Type 1 error α, the phase-2 upper control limit UCL_2 can be calculated by

$$\mathrm{UCL}_2 = \frac{p(m-1)(m+1)}{m(m-p)} F_\alpha(p, m-p) \qquad (11)$$

Based on this control limit, a process fault can be detected if statistic T^2 of a waveform signal exceeds the control limit UCL_2; otherwise, the process is concluded under a normal working condition.

For a new process, the in-control data identified at phase 2 can be added into the preliminary in-control data identified at phase 1. Increasing the sample size can improve the estimation accuracy of $\bar{\mathbf{x}}$ and $\mathbf{S}$ at phase 1. This iterative process monitoring while simultaneously updating the estimate of class parameters can continuously improve the monitoring performance.

3.2.2. *New fault identification*

In order to accumulate knowledge of process faults, new fault classes should be formed and continuously accumulated whenever a new process fault is identified. In this study, fault waveform signals are assumed to follow a p-dimensional multivariate normal distribution. $\mu(k)$ and $\Sigma(k)$ represent parameters of fault cluster k. In a fault cluster, the number of fault samples and the dimension of variables will influence the accuracy of parameter estimates. Therefore, increasing the sample size in a fault cluster, through continuously accumulating fault samples during production monitoring, can improve the estimation accuracy. Moreover, the reduction of variable dimension via feature subset selection can also improve the estimation and diagnosis accuracy.

After initially detecting a fault using UCL_2 in step 1 as shown in Fig. 3, a test will be conducted to determine whether this fault belongs to a known fault cluster or a new fault cluster. Two step analyses, called fault information assessment, are proposed for this purpose.

Step 1: Identify the closest cluster for a newly detected fault

When a process fault is detected, a multiclass classification will be conducted to identify the known fault cluster, to which the detected fault is most likely to belong. This identified cluster is called the closest cluster and noted as cluster k^*. A piecewise classifier is designed for this task, and will be discussed in detail in next subsection.

Step 2: Perform acceptance testing for the identified cluster

For cluster k^*, the mean and variance are calculated using its samples. Then, Hotelling T^2 control limits respective to phase 1 and phase 2 are calculated using Equations 10 and 11. The control chart at phase 2 is used to test whether the newly detected fault belongs to cluster k^* under a given Type 1 error. If the control chart indicates that T^2 of the new detected fault is within the control limit, it is concluded that the newly detected fault belongs to cluster k^*; otherwise, the newly detected fault is concluded to be a ''new'' fault (i.e., not in the current known fault clusters). Thus, a new fault cluster needs to be formed.

It should be emphasized that only a selected subset of features are used to calculate the T^2 control limit in the fault information assessment; this is different from previous normal working condition monitoring where all features are considered. The main reason of selecting a subset of features is that the number of fault samples is generally very small at the beginning of monitoring system development, and less than the number of all features. In this situation, the estimation of fault cluster parameters is an ill-posed problem. Thus, it is impossible to obtain an accurate estimate for the cluster parameters if the number of features used is larger than the number of samples. As a result, the method of how to select the subset of features needs to be investigated to represent the fault characteristics in the fault information assessment. The optimal selection of the feature subset will be discussed in Section 3.4.

3.3. *Piecewise classifiers*

When a fault is detected, fault classification will be performed to identify the closest known fault cluster. In this study, it is assumed that the process fault is mainly due to the mean shift, thus, a linear piecewise linear classifier is recommended (Fukunaga, 1990):

$$\text{Class } k^* = \max_k \left\{ \hat{\mu}(k)^T \hat{\Sigma}^{-1} \mathbf{x}_i - \frac{1}{2} \hat{\mu}(k)^T \Sigma^{-1} \hat{\mu}(k) + \ln P_k \right\} \tag{12}$$

where P_k is the prior probability of fault cluster k. It can be assumed that P_k is the same for all clusters if there is no prior knowledge available. The closest cluster k^* is identified from all known fault classes. $\hat{\Sigma}$ is considered as a pooling variance of all fault samples, and is estimated by

$$\hat{\Sigma} = \frac{1}{\left(\sum_k n_k\right) - 1} \sum_k \sum_i (\mathbf{x}_{i|k} - \hat{\boldsymbol{\mu}}(k)).(\mathbf{x}_{i|k} - \hat{\boldsymbol{\mu}}(k))^T \tag{13}$$

where $x_{i|k}$ is sample i in cluster k, and n_k is the number of samples in cluster k.

3.4. *Optimal feature subset selection*

3.4.1. *Optimal feature selection under a given dimension of subset features*

The reduction of feature space dimension can avoid an ill-posed decision-making problem when the number of fault samples is less than the dimension of a feature space. A subset of features is selected from the whole feature space for this purpose. Under the given dimension of a feature subset, the optimal features are selected based on a criterion of class separability defined as Fukunaga (1990):

$$J = \text{trace}(\mathbf{S}_m^{-1} \mathbf{S}_b) \tag{14}$$

where $\mathbf{S}_m = E[(\mathbf{x} - \hat{\boldsymbol{\mu}})(\mathbf{x} - \hat{\boldsymbol{\mu}})^T]$ is a mixture scatter matrix representing the covariance matrix of all samples regardless of their class assigments; $\hat{\boldsymbol{\mu}}$ is the estimated mean of all fault samples. $\mathbf{S}_b = \sum_k P_k(\hat{\boldsymbol{\mu}}(k) - \hat{\boldsymbol{\mu}})(\hat{\boldsymbol{\mu}}(k) - \hat{\boldsymbol{\mu}})^T$ is a between-class scatter matrix representing the scatter of the estimated cluster mean around the estimated mean of all samples. $\hat{\boldsymbol{\mu}}(k)$ is equal to the estimated mean of cluster k. The optimal feature subset is selected for a given dimension of a feature subset by maximizing J from all combinations of feature subsets.

3.4.2. *Optimal dimension of a feature subset in terms of classification errors*

For a multiclass classification problem, the classification error of class k is defined by

$$\begin{aligned} \beta_k &= \sum_{j \neq k} \text{prob}[D_j | H_k] \\ &= \sum_{j \neq k} \text{prob}[\mathbf{x}_i - \hat{\boldsymbol{\mu}}_j)^T \mathbf{S}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_j) < \text{UCL}_2(j)] \\ &= \sum_{j \neq k} \text{prob}[(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k + \Delta\boldsymbol{\mu})^T \mathbf{S}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k + \Delta\boldsymbol{\mu}) \\ &< \text{UCL}_2(j)] \end{aligned} \tag{15}$$

where prob $D_j|H_k$ means the probability of the classified class for data $\mathbf{x}_i$ is class j but its true class is class k, which can also be noted as $\beta_{j|k}$. The $\text{UCL}_2(j)$ is the phase-2 control limit, which is constructed based on the available fault samples in class j. $\Delta\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}_k - \hat{\boldsymbol{\mu}}_j$ is the mean difference between class k and class j. It is known that $\delta\ (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k + \Delta\boldsymbol{\mu})^T \mathbf{S}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k + \Delta\boldsymbol{\mu})$ follows a noncentral F distribution (Pearson and Hartley, 1972).

In terms of the classicification performance, the optimal number of features in a feature subset needs to be considered because it has two opposing effects on error β_k:

(1) The decrease of the number of features p can monotonically decrease δ. A smaller δ results in a smaller change of statistic T^2, which has a tendency to increase the error $\beta_{j|k}$ for a noncentral F statistic.

(2) For ULC_2 in Equation 11, it can be easily seen that for a given m, the constant multiplier of $p(m-1)(m+1)/[m(m-p)]$ is proportional to $p/(m-p) = -1 + m/(m-p)$. Thus, this constant multiplier is monotonically decrease as the value of p decreasing under the condition of $m > p$. Moreover, $F_\alpha(p, m-p)$ also decreases with a decreased value of p. As a result, UCL_2 will be tighter with a decreased value of p, which will have a tendency to reduce the error of $\beta_{j|k}$.

Considering the above two opposite tendencies on error β_k, the optimal dimension of feature subsets needs to be considered to obtain a minimum classification error.

3.5. *Classification error estimation*

It is known that the number of samples can affect the estimation of a classification error. The bias of the classification error comes entirely from a finite design sample set which is used to estimate the cluster parameters. The variance comes predominantly from a finite test sample set that is used to test the classification performance.

The cross-validation method (Fukunaga, 1990) can provide an upper boundary of the Bayes error. In addition, it can provide a better performance than the holdout method when the number of samples is limited. In the cross-validation method, one sample is excluded from a cluster. Then, the respective classifier is redesigned based on the remaining $n_k - 1$ samples, which is used for testing this excluded sample. This procedure is repeated n_k times for testing all n_k samples, and the number of misclassified samples is counted in order to obtain the estimate of the classification error. The expectation of this error gives the upper boundary of the Bayes error. Obviously, the advantage of the cross-validation method, also called leave-one-out method, is that it fully utilizes all sample information for the classifier performance evaluation.

In order to reduce classifier computation time when a sample is excluded, the re-estimation of parameters can be conducted simply by using the following perturbation equations:

$$\hat{\boldsymbol{\mu}}_i(k) = \hat{\boldsymbol{\mu}}(k) - \frac{1}{n_k - 1}[\mathbf{x}_i - \hat{\boldsymbol{\mu}}(k)] \tag{16}$$

$$\hat{\Sigma}_j k = \hat{\Sigma}(k) + \frac{1}{n_k - 2}\hat{\Sigma}(k) - \frac{n_k}{(n_k - 1)(n_k - 2)}[\mathbf{x}_i - \hat{\boldsymbol{\mu}}(k)][\mathbf{x}_i - \hat{\boldsymbol{\mu}}(k)]^T \tag{17}$$

where $\hat{\boldsymbol{\mu}}_i(k)$ and $\hat{\sum}_i(k)$ are the updated estimates of mean and variance for classifier k when sample $\mathbf{x}_i$ is excluded.

This classification error is used as an assessment of the diagnostic performance of a system, which will also serve as the feedback to determine the optimal dimension of a feature subset as discussed in Section 3.4.2.

4. Example: Automatic feature extraction from tonnage signals in a stamping process

4.1. *Process description*

A tonnage signal representing a stamping force has played an important role in stamping process monitoring and fault diagnosis. Analysis of waveform changes of a tonnage signal can assess the stamping process conditions and also has the potential to predict part quality. Generally, in order to reduce process variations in production, a consistent tonnage waveform signal is required; it is expected to be the same as that found under the normal working conditions. If a tonnage signal is detected beyond the boundary of the normal working condition, a process fault is detected.

Figure 4 shows three typical process variables in a stamping press, which are blank thickness, shut height, and punch speed. The changes of those process variables are considered as process faults in the paper. The tonnage sensors mounted on the press uprights are used to measure the tonnage force. Figure 5 shows the major portion of one cycle of a tonnage signal in a double-action process. The press crank angle in the X axis is generally used as a reference to divide one cycle of operation into different working stages (Jin and Shi, 1999). The inner and outer tonnage signals are measured by the same strain gage sensors mounted on the press columns. The inner tonnage signal, which corresponds to the crank angle region of [159.29°, 184.78°] as shown in Fig. 5, is used to demonstrate the developed methodology. In

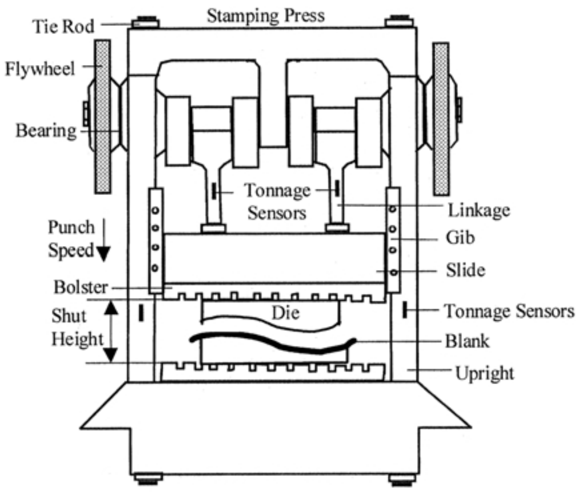

Fig. 4. A stamping press and process variables.

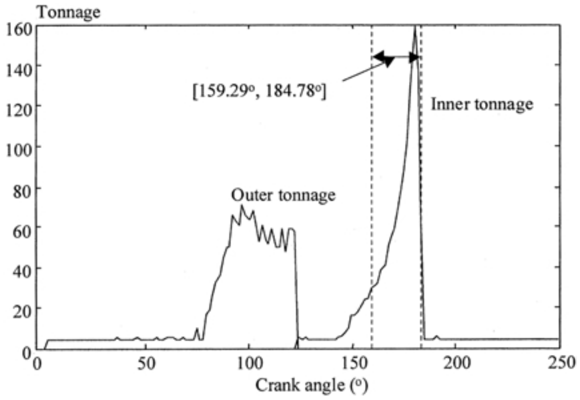

Fig. 5. One cycle of a tonnage signal.

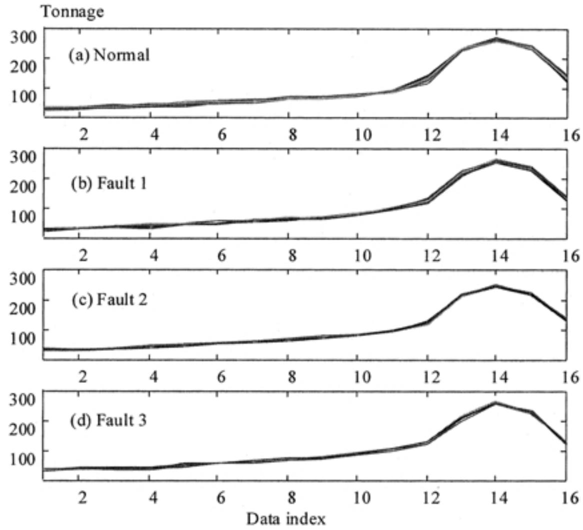

Fig. 6. Tonnage signals under different conditions.

this segment, 16 data points are included to satisfy the requirement of the Haar transform.

Four sets of data, as shown in Fig. 6, are collected under the different working conditions in production. For the convenience of using the Haar transform, the data index is used as the X axis in Fig. 6; it has a linear relationship with the press crank angle. Figure 6(a) shows one set of normal working condition data with 27 samples, and Fig. 6(b)–(d) show three different fault condition data with nine samples in each set. All three-fault conditions have a decreased inner shut height (from the normal 95.9440 to a faulty 95.9435 inches). Fault 1 and fault 3 have an additional decrease on the outer shut height (from the normal 83.6650 to a faulty 83.6553 inches), and fault 3 uses a thicker blank sheet metal. Fault 2 has an increased outer shut height (from the normal 83.6650 to a faulty 83.6725 inches), a decreased punch speed (from the normal 14 to a faulty 7 stroke/min), and a thicker blank. The range of each variable is selected so that the variables beyond those extreme conditions will affect the part quality. A summary of these different process setups with the coded values of the variables is shown in Table 1, where ''0'' corresponds to the normal setup value, and '' – 1'' and ''1'' correspond to the lower and higher setup values (i.e., faults), respectively. From Fig. 6, it can be seen that all tonnage signals have good repeatability under each working condition. Figure 7 provides a comparison of the means of tonnage profiles under different work conditions. From Fig. 7, it can be seen that although the profile changes of a tonnage signal are good indicators for process variable changes, it is very difficult to classify them by using a trial-and-error method due to their complexity of the profile changes. The proposed method is used to analyze the aforementioned data to demonstrate the effectiveness and procedures of the proposed methodology.

In the analysis, we use three-fault data and normal-condition data to simulate the real production environment by considering the incoming faults sequentially. At the beginning, only normal production data is available. Based on this, normal production condition monitoring is performed by developing the phase 1 control limit using the normal production data, i.e., the preliminary data. This result will then lead to the design of a phase 2 control limit as discussed in Section 3.2.1. After that, this phase 2

Table 1. Process setups under different working conditions

Variable	*Blank thickness*	*Outer shut height*	*Inner shut height*	*Punch speed*
Normal condition	0	0	0	0
Fault 1	0	– 1	– 1	0
Fault 2	1	1	– 1	– 1
Fault 3	1	– 1	– 1	0

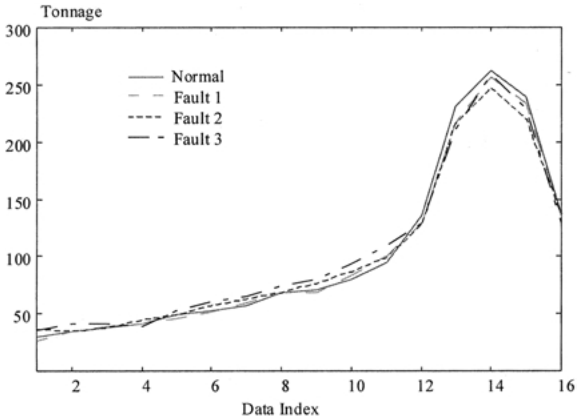

Fig. 7. Comparison of mean profiles of different conditions.

control limit is used to monitor the production data. In production, it is assumed that fault 1 occurs first and generates nine fault samples. The fault is detected by the phase 2 control limit. Since only one fault is available, there is no need for a fault classification study in this stage. During continuous production, fault 2 occurs and the fault signal is detected by the phase 2 control limit. At this stage, an analysis is needed to determine if this newly detected fault belongs to the known fault (i.e., fault 1). If it does, this newly detected fault sample will be added to the known fault cluster, and the statistic parameters of this fault cluster will be updated for more accurate parameter estimation. If it does not, as is the case in this example, a new fault cluster has to be formed and an optimal subset of features will be selected to represent those two fault characteristics. A piecewise linear classifier will then be designed using the optimal feature subset for future fault classification. As production continuous, fault 3 occurs and is detected. After that, the above procedures are repeated. If the fault does not belong to those two known faults, a situation that is true in this example, the new feature subset to maximize the separability of those faults will be optimized again. Then a piecewise linear classifier will be redesigned.

4.2. *Analysis results*

Based on the procedures shown in Fig. 3, the Haar transform is used first for feature extraction. Five Haar coefficients $c_0^0\, c_1^1\, c_2^2\, c_3^4\, c_4^8$ are kept after data denoizing using a universal threshold (Donoho and Johnstone, 1994), which are considered as the whole feature space. The optimal feature subset will be formed by selecting optimal features from these five features.

The T^2 control limits are established based on these 27 samples of normal working conditions using Equations 10 and 11 respective to phase 1 and phase 2, where $p = 5$ corresponds to five Haar coefficients, $m = 27$ represents the number of samples, and $\alpha = 0.05$ respective to the Type 1 error. Therefore, in normal working condition monitoring, phase 1 control limit $\mathrm{ULC}_1 = 15.14$ is used to eliminate outliers from those preliminary data. In this example, there is no outlier found in this step. The T^2 control limit in phase 2, $\mathrm{UCL}_2 = 16.31$, is used to check whether the production data is in normal working conditions.

As the process faults occurred sequentially, the fault signals were detected using the phase 2 control limit. In the following discussion, we discuss the analysis only after fault 2 occurs in the production. In this case, more than two types of faults are found. Thus, an optimal selection of fault feature subset will be conducted by maximizing the separability of these faults as discussed in Section 3.4. Table 2 gives the selected optimal Haar coefficients under the different dimensions of the feature subsets.

After selecting the feature subset, a piecewise linear classifier given in Equation 12 and UCL_2 control limit given in Equation 11 is used for fault sample classification, where p is the different dimensions of the feature subset, and $m = 9$. The cross-validation approach is used to estimate the classification error as discussed in Section 3.5, and is given by the percentage value in the bracket in Table 2. The optimal dimension of the feature subset is determined by the minimum classification error as discussed in Section 3.4.2. In this case study, the dimension of

Table 2. Optimal feature subset selection

	Optimal features (percentage of classification errors)				
Subset dimension	1	2	3	4	5
Faults 1, 2	$c_4^8(0\%)$	$c_0^0, c_1^1(0\%)$	$c_0^0, c_1^1, c_4^8(0\%)$	$c_0^0, c_1^1, c_2^2, c_4^8(0\%)$	(0%)
Faults 1, 2, 3	$c_2^2(33.3\%)$	$c_0^0, c_1^1(55.5\%)$	$c_0^0, c_1^1, c_4^8(0\%)$	$c_0^0, c_1^1, c_3^4, c_4^8(0\%)$	(0%)

these optimal feature subsets is equal to 1 and 3 corresponding to the two faults and three faults, respectively. If the classification error is equal, the smaller dimension should be selected as the case when the first two faults are analyzed.

5. Conclusion

In this paper, a new methodology is developed for developing a diagnostic system by using waveform signals with limited or with no prior fault information. In this research, the focus is on automatic feature extraction and optimal feature subset selection during the use of a diagnostic system in a manufacturing process. This research result will accomplish in-process and continuous diagnostic performance enhancement and improvement. A real case study of stamping tonnage signal analysis has been used to demonstrate the effectiveness of the proposed methodology.

It should be pointed out that the significance of the work is not limited to a new diagnostic system development where historical faults are not available and where the understanding of a process is limited. The developed methodology can also be used in an existing diagnostic system where continuous diagnostic performance enhancement and improvement are desired. In addition, the developed method has potential to be used in data mining and knowledge discovery when massive waveform data are available but limited knowledge about the data pattern exists. In this situation, the automatic feature extraction and classifications can be used to group the massive data into organized clusters for further analysis. The authors of the paper have been devoting their efforts to this study.

There are a few issues to be studied further in the methodology development. Examples of those open issues include: (1) Theoretical assessment of the impact of Type 1 and Type 2 error of the monitoring algorithms on the performance of the diagnostic system; (2) the impact of the sample size on the diagnostic performance; and (3) the upper bound of the diagnostic performance improvement when implementing the proposed methodology.

Appendix

For a statistic $Q = m\mathbf{y}^T\mathbf{W}^{-1}\mathbf{y}$, where $\mathbf{y} \sim N_d(\mathbf{0}, \Sigma)$, $\mathbf{W} \sim W_d(m, \Sigma)$, and $\mathbf{y}$ and $\mathbf{W}$ are statistically independent. N_d and W_d are used to denote a d-dimensional normal and Wishart distribution respectively. Then, it can be obtained as follows (Seber, 1984, pp. 30–31):

$$\frac{m-d+1}{d}\frac{Q}{m} \sim F(d, m-d+1) \qquad \text{(A-1)}$$

At phase 1, the preliminary observations are denoted $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$, where each $\mathbf{x}_i$ is a vector with a dimension of p, and $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \Sigma)$. Therefore, the estimates of the mean vector and variance have the following distributions:

$$\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, \Sigma/m); \quad \mathbf{W} = (m-1)\mathbf{S} \sim W_p(m-1, \Sigma) \qquad \text{(A-2)}$$

Because $\mathbf{x}_i$ is independent of other observations $\mathbf{x}_j (j \neq i)$, so the dependency of $\bar{\mathbf{x}}$ on $\mathbf{x}_i$ can be decomposed by

$$\mathbf{x}_i - \bar{\mathbf{x}} = \mathbf{x}_i\left(1 - \frac{1}{m}\right) - \sum_{j\neq i}\frac{\mathbf{x}_j}{m} = \frac{m-1}{m}\mathbf{x}_i - \frac{1}{m}\sum_{j\neq i}\mathbf{x}_j \qquad \text{(A-3)}$$

Thus, the variance of $\mathbf{x}_i - \bar{\mathbf{x}}$ can be calculated by

$$\mathrm{Var}(\mathbf{x}_i - \bar{\mathbf{x}}) = \left(\frac{m-1}{m}\right)^2\Sigma + \frac{m-1}{m^2}\Sigma = \frac{m-1}{m}\Sigma \qquad \text{(A-4)}$$

Equivalently, it can be denoted as

$$\mathbf{y} = \sqrt{\frac{m}{m-1}}(\mathbf{x}_i - \bar{\mathbf{x}}) \sim N_p(0, \Sigma) \qquad \text{(A-5)}$$

If a statistic G is defined as

$$G = (m-1)\mathbf{y}^T\mathbf{W}^{-1}\mathbf{y} = \left(\frac{m}{m-1}\right)(\mathbf{x}_i - \bar{\mathbf{x}})^T\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) \qquad \text{(A-6)}$$

thus,

$$\frac{m-1-p+1}{p}\frac{G}{m-1} \sim F(p, m-1-p+1) \qquad \text{(A-7)}$$

Substitute Equation A-6 into Equation A-7, the statistic $T^2 = (\mathbf{x}_i - \bar{\mathbf{x}})^T\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$ (satisfies:

$$T^2 \sim \frac{p(m-1)^2}{m(m-p)}F(p, m-p) \qquad \text{(A-8)}$$

Acknowledgment

This work was supported by the NSF CAREER award: DMI 9624402 and NSF Grant: DMI 9713654.

References

Alt, F. B. (1985) Encyclopedia of Statistical Science, Vol. 6, N. L. Johnson and S. Kotz, (eds.) John Wiley, New York, pp. 110–122.

Coifman, R. R. and Yale, F. M. (1992) Adapted waveform analysis and denoising, Meyer, Y. and Roques, S. (eds.) *Progress in Wavelet Analysis and Applications*, in the *Proceeding of the International Conference*, *Wavelets and Applications*, Toulouse, France, pp. 63–93.

Deibert, R., and Hofling, T. (1992) Model based fault diagnosis of industrial robot, a comparison of different methods, Judd, R. P. and Kheir, N. A., (eds.), IFAC Workshop: Intelligent Manufacturing Systems, Dearborn, MI.

Donoho, D. L. and Johnstone, I. M. (1994) Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81**, 425–455.

Daubechies, I. (1992) *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Fukunaga, K. (1990) *Introduction to Statistical Pattern Recognition*, Academic Press, New York.

Hotelling, H. (1947) Multivariate quality control. *Techniques of Statistical Analysis*, Eisenhart, Hastay and Wallis (eds.), McGraw-Hill, New York.

Jin, J. and Shi, J. (1999) Feature-preserving data compression of stamping tonnage information using wavelets. *Technometrics*, **41**(4), 327–339.

Kasashima, N., Mori, K., Ruiz., G. H. and Taniguchi, N. (1995) Online failure detection in face milling using discrete wavelet transform. *Annals of the CIRP*, **44**(1), 483–487.

Kharin, Y. (1992), *Robustness in Statistical Pattern Recognition*, Kluwer Academic Publishers, Dordrecht.

Mori, K., Kasashima, N., Yoshioka, T. and Ueno, Y. (1996) Prediction of spalling on a ball bearing by applying the discrete wavelet transform to vibration signals. *Wear 195, Elsevier Science S. A.*, pp. 162–168.

Pearson, E. S. and Hartley, H. O. (1972) *Biometrika Tables for Statisticians*, Cambridge. Published for the Biometrika Trustees at the University Press, **II**, pp. 66–68.

Seber, G. A. F. (1984) *Multivariate Observations*, John Wiley & Sons, New York, NY.

Strang, G. and Nguyen, T. (1996) *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA.

Tansel, I. N., Mekdeci, C., Rodriguez, O. and Uragun, B. (1993) Monitoring drill condition with wavelet based encoding and neural networks. *International Journal of Machine Tools Manufacturing*, **33**(4), 559–575.

Tracy, N. D., Young, J. C. and Mason, R. L. (1992) Multivariate control charts for individual observations. *Journal of Quality Technology*, **24**(2), 88–95.