Multisensory Fusion based Virtual Tool Wear Sensing for Ubiquitous Manufacturing

Jinjiang Wang[1], Junyao Xie[1], Rui Zhao[2], Laibin Zhang[1], and Lixiang Duan[1,*]

[1] School of Mechanical and Transportation Engineering
China University of Petroleum, Beijing, China, 102249

[2] School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore, 639798

* Corresponding author: duanlx@cup.edu.cn

Abstract

Pervasiveness of ubiquitous computing advances the manufacturing scheme into a ubiquitous manufacturing era which poses significant challenges on sensing technology and system reliability. To improve manufacturing system reliability, this paper presents a new virtual tool wear sensing technique based on multisensory data fusion and artificial intelligence model for tool condition monitoring. It infers the difficult-to-measure tool wear parameters (e.g. tool wear width) by fusing in-process multisensory data (e.g. force, vibration, etc.) with dimension reduction technique and support vector regression model. Different state-of-the-art dimension reduction techniques including kernel principal component analysis, locally linear embedding, isometric feature mapping, and minimum redundancy maximum relevant method have been investigated for feature fusion in a virtual sensing model, and the kernel principal component analysis performs best in terms of sensing accuracy. The effectiveness of the developed virtual tool wear sensing technique is experimentally validated in a set of machining tool run-to-failure tests on a computer numerical control milling machine. The results show that the estimated tool wear width through virtual sensing is comparable to that measured offline by a microscope instrument in terms of accuracy, moreover, in a more cost-effective manner.

Keywords: Tool wear estimation, virtual sensing, feature fusion

1 INTRODUCTION

Ubiquitous computing enables a variety of information technologies and communication services connect with the distributed computational resources via internet [1]. With the advancement of ubiquitous computing and radio frequency identification technology (RFID), it promotes product design and manufacturing into a new paradigm, named UbiDM (design and manufacture via ubiquitous computing technology) [2, 3]. It changes the manufacturing scheme from mass-production/consumption to a new customized and sustainable manufacturing mode. To achieve this, advanced sensing techniques and computing intelligence are needed to improve system flexibility and inventory turnover, minimise wastage, improve product quality and enhance on-time delivery [4, 5]. However, such benefits highly rely on the reliability of a manufacturing system, thus system reliability becomes a crucial important aspect in ubiquitous manufacturing.

Machining tool is a major element in a manufacturing system and its failure (e.g. typically tool wear and breakage) can attribute up to 20% of machine downtime [6]. To enhance the system reliability, much research effort has been put on machining tool condition monitoring which mainly incorporates sensing data acquisition, signal denoising and processing, feature extraction and selection, fault diagnosis and prognosis, and maintenance decision making [7]. Increasing demand for system reliability has accelerated the integration of sensors into manufacturing system for timely acquisition of working status of machinery. With the advancement of ubiquitous computing, ubiquitous sensing emerges as an active research area. It has been investigated in real-time shop-floor scheduling [8], environment monitoring [9], electrical household appliance [10], and human healthcare [11].

In the context of tool condition monitoring, a variety of sensing techniques have been instrumented to acquire machining tool conditions. According to the correlation between sensing

parameters and tool conditions [7], these sensing techniques can be categorized into direct sensing and indirect sensing methods. Direct sensing techniques measure actual quantities that directly indicate tool conditions, e.g. tool wear width. Traditionally, tool wear was measured using a tool-maker's microscope under laboratory conditions. This requires a human inspector to determine the worn region based on the textural difference between the worn and unworn surfaces [12]. In-situ direct sensing techniques, such as CCD cameras, radioactive isotopes, laser beams, and electrical resistances, have been investigated with high degree of accuracy in laboratory settings to observe fundamental measurable phenomena during machining processes [13-15]. However, such direct sensing techniques usually involve high cost, and present some practical limitations caused by accessing problems during machining, inference with chips, and the usage of cutting fluid. Thus they are mainly for intermittent operations.

On the contrary, indirect sensing techniques measure the auxiliary in-process quantities (e.g., force, vibration, and acoustic emission, etc.) that indirectly indicate tool conditions. Tool wear causes the increases of friction and heat generation, thus consequently causes the changes of in-process parameters, such as cutting force [16], vibration [17], acoustic emission [18], strain [19], eddy-current displacement [20], and spindle motor current [21], etc. Comparing to direct sensing, indirect sensing methods are less costly and enables continuous detection of all changes (e.g. tool breakage, tool wear, etc.) to signal measurements. The Pros and Cons of direct sensing and indirect sensing methods are summarized in Table 1.

To sum up, direct sensing measures direct indicators of tool conditions, but it is usually performed offline and thus interrupts normal machine operations. On the other hand, indirect sensing can continuously measure in-process parameters, but the obtained information is indirect indicators of tool conditions. To bridge the gap between direct sensing and indirect sensing, virtual sensing, as a complement to physical sensing, has emerged as a viable, noninvasive, and cost effective method to infer difficult-to-measure or expensive-to-measure parameters in

dynamic systems based on computational models [22]. It has been investigated for active noise and vibration control [23], industrial process control [24], building operation optimization [25], lead-through robot programming [26], product quality of semiconductor industry [27], and tool condition monitoring [28, 29].

Table 1. Comparison between direct sensing and indirect sensing techniques.

Category	Sensing techniques	Pros	Cons
Direct sensing	Microscope, CCD camera, Electrical resistances, Radioactive isotopes	Accurate, direct indicators of tool conditions	High cost, limited by operating environment, mainly for offline or intermittent monitoring
Indirect sensing	Cutting force, vibration, sound, acoustic emission, temperature, spindle power, displacement	Less complex, low cost, suitable for continuous monitoring in practical applications	Indirect indicators of tool conditions

There is an extensive literature on developing virtual sensing models with a focus of artificial intelligence models. In [28], an artificial neural network model is investigated to infer the state of insert wear from translational vibration measurements on a milling machine. Bayesian network is studied for tool breakage detection utilizing the in-process electrical power signal [29]. A classifying artificial neural network ensemble approach is investigated to estimate simulation workload in cloud manufacturing [30]. Given the high cost and practical constraints to obtain data samples in practice, support vector regression with good generalization capability attracts much research interest. It requires less data samples comparing with artificial neural network, and has been investigated for tool wear estimation in [31-33].

Due to the interference of complex operating conditions and the limited applicability of a single sensor, multiple modalities of sensors have been instrumented to measure different aspects of tool conditions. However, the increased amount of data samples inevitably brings data redundancy and model overfitting problems. To address these issues, this paper presents a multisensory fusion based virtual tool wear sensing method on a support vector regression basis.

Different dimension reduction methods including kernel principal component analysis (KPCA), locally linear embedding (LLE), isometric feature mapping (ISOMAP), minimum redundancy maximum relevance (mRMR) have been investigated for feature selection and fusion. The fused features from force and vibration in-process measurements are then fed into support vector regression model to infer the actual quantities of tool conditions. The performance of different feature fusion methods is compared using experimental studies on a computer numerical control (CNC) milling machine.

The main contribution of this study rests on the following: 1) a multisensory fusion based virtual tool wear sensing framework is presented to bridge the gap between direct sensing and indirect sensing methods; and 2) different dimension reduction techniques are evaluated for virtual tool wear sensing, and the technique of KPCA with the best performance is identified by an experimental study. The rest of the paper is constructed as follows. After introducing the theoretical background of sensing fusion techniques and support vector regression in Section 2, details of the multisensory fusion based virtual sensing method is discussed in Section 3. The effectiveness of the presented technique is experimentally demonstrated in Section 4 based on direct and indirect sensing data acquired using a ball nose tungsten carbide cutter on a CNC milling machine. Finally, conclusions are drawn in Section 5.

2 THEORETICAL FRAMEWORK

2.1 Data Fusion Techniques

2.1.1 Kernel Principal Component Analysis

Kernel principal component analysis (KPCA) is a nonlinear version of principal component analysis (PCA) and has been widely used for feature selection and fusion applications. The key idea of KPCA is to define a nonlinear mapping function $\phi(\bullet)$ which transforms the sample data into a high-dimensional data space, and the transformed sample data is then analyzed using

traditional principal component analysis [34]. It transforms a set of observations of possible correlated variables into a set of uncorrelated variables called principal components. The first principal component has the largest variance, and each succeeding principal component has comparative lower variance orthogonal to the preceding principal components. The first several principal components can represent the original data with minimal mean squared approximation error, and thus KPCA can be used in dimensionality reduction.

Mathematically, given a set of input vectors $(X_i(1), X_i(2), \ldots, X_i(\mathrm{m}))^T$, $i = 1, 2, \ldots, p$, the sample data X_i is mapped into $\phi(X_i)$ via the nonlinear kernel function $\phi(\bullet)$, i.e. $X_i \to \phi(X_i)$. With the assumption of centered data $\frac{1}{p}\sum_{i=1}^{p}\phi(X_i)=0$, the principal components are obtained by solving eigenvalue problem in KPCA.

$$\lambda_i u_i = C u_i \tag{1}$$

where C is the sample covariance matrix of $\phi(X_i)$, λ_i is one of the eigenvalues of covariance matrix C, and u_i is the corresponding eigenvector. The covariance matrix is constructed as:

$$C = \frac{1}{p}\sum_{i=1}^{p}\phi(X_i)\phi(X_i)^T \tag{2}$$

Define a Gram matrix K with its elements as

$$k_{ij} = \phi(X_i)^T\phi(X_j) = \phi(X_i)\bullet\phi(X_j) \tag{3}$$

where x_i and x_j are the sample vectors. Assuming $k(\bullet)$ is a symmetric kernel function, the dot production in Eq. (3) can be replaced by a kernel function $k(\bullet)$ based on the Mercer's theorem. Since the data points need to be centered in the feature space, the centered kernel matrix $\tilde{K}$ is defined as [35]:

$$\tilde{K} = K - A_p K - K A_p + A_p K A_p \tag{4}$$

where A_p is a $p \times p$ matrix with $(A_p)_{ij} = 1/p$. The eigenvalue equation (1) can be rewritten as

$$p\lambda_i u_i = \tilde{K} u_i \tag{5}$$

Then the n_{th} kernel principal component is readily obtained by projecting the observations in the direction of the n_{th} eigenvector [34].

$$u_n \phi(X_i) = \sum_{j=1}^{p} u_n k(X_j, X_i), \; i = 1, 2, \ldots, p \tag{6}$$

Since the number of eigenvectors is the same as the sample size in KPCA, it can deal with nonlinear problems which cannot be solved by PCA. By calculating the accumulated contribution rate (e.g., 95%), the number of the most significant principal components can be selected for dimensionality reduction.

$$\sum_{k=1}^{q} \lambda_k / \sum_{k=1}^{p} \lambda_k \geq 95\% \tag{7}$$

2.1.2 Locally Linear Embedding Algorithm

Locally linear embedding (LLE) is a nonlinear dimension reduction method by computing the low dimensional, neighborhood preserving embedding of high dimensional data. It attempts to discover the underlying nonlinear structure (nonlinear manifold) in high dimensional data by exploiting the local symmetries of linear reconstructions [36].

Suppose that the data consists of N real-valued vectors X_i which are sampled from an underlying manifold as illustrated in Fig. 1. With sufficient data samples (such that the manifold is well-sampled), each data point and its neighbors are expected to lie on or close to a locally linear patch of the manifold. Thus, the local geometry of these patches is characterized by linear coefficients and each data point can be reconstructed.

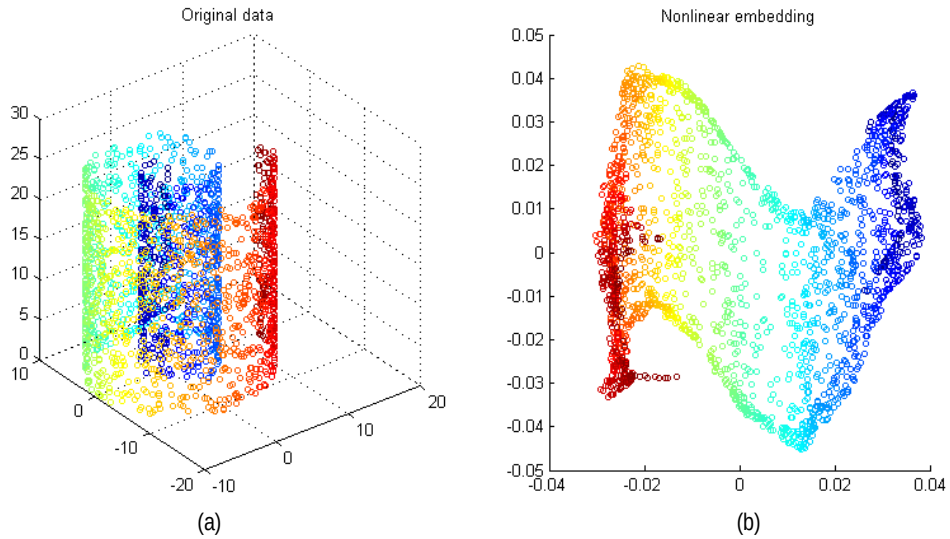

Fig. 1. Illustration of locally linear embedding algorithm, a) data points sampled from underlying 3-dimensional manifold, b) 2-dimensional neighborhood-preserving mapping by LLE.

First, the number of the nearest neighbors per data point is identified by calculating the Euclidean distances between neighbor points and the data point of interest. Next, the weights matrix is computed by minimizing the reconstruction error of the data point from its neighbors. The cost function is formulated by adding up the squared distances between all data points and their corresponding constructions as [36]:

$$\varepsilon(\mathrm{W}) = \sum_i \left| X_i - \sum\nolimits_j W_{ij} X_j \right|^2, \quad \sum\nolimits_j W_{ij} = 1 \tag{8}$$

where W_{ij} represents the contribution weights of j_{th} neighbor data points X_j to the i_{th} reconstructed data point X_i. The weights W_{ij} are optimized by solving a least squares problem. Finally, each high dimensional observation X_i is mapped to a low dimensional vector Y_i representing the global internal coordinates on the manifold. This is achieved by selecting a d-dimensional vector Y_i to minimizing the cost function as [36]:

$$\Phi(Y) = \sum_i \left| Y_i - \sum_j W_{ij} Y_j \right|^2 \tag{9}$$

It is defined as a quadratic form of vector Y_i. It is constructed similarly to Eq. (8). The difference is that the weights W_{ij} is fixed in Eq. (9) to optimize the coordinates Y_i. The cost function is minimized by solving an eigenvector problem subject to constraints that make the problem well-posed. Therefore, it is pretty straightforward to implement the locally linear embedding algorithm as only one free parameter needs to be optimized. The number of neighbors per data point K is determined firstly. Once the neighbors are chosen, the optimal weights W_{ij} and coordinates Y_i are computed by standard methods in linear algebra.

2.1.3 ISOMAP

As a representative manifold learning technique, ISOMAP is commonly used for nonlinear dimensionality reduction by mapping high-dimensional data to a lower dimensional space. ISOMAP extends metric multidimensional scaling (MDS) by incorporating the geodesic distances instead of pairwise Euclidean distances in the computation of graph shortest path distances [37]. The mathematical formulation of the ISOMAP algorithm is briefly discussed as follows.

First, a neighborhood graph G is constructed by determining which points are neighbors on the manifold M based on the distances d_X(i,j) between pairs of points i, j in the input space X. Two simple methods of ε-ISOMAP and K-ISOMAP can be used to determine the neighborhood points. The data points j within a fixed radius ε of point i are considered as neighbors in ε-ISOMAP while the K nearest neighbors points j are selected in K-ISOMAP.

Next, the geodesic distances $d_M(i,j)$ between all pairs of points on the manifold M are estimated by computing the shortest path distances $d_G(i,j)$ in the graph G.

$$d_G(i,j) = d_G(i,k) + d_G(k,j),\ k = 1, 2, \ldots N \tag{10}$$

This could be calculated by Dijkstra's algorithm $\left(o\left(kn^2 \log n\right)\right)$ and Floyd's algorithm $\left(o\left(n^3\right)\right)$. A set of data pair path distances $d_G(i,j)$ are calculated and used to construct graph distances matrix $D_G=\{d_G(i,j)\}$, which contains the shortest path distances between all pairs of points in graph G.

Finally, the classical MDS is applied to the matrix D_G, constructing an embedding of the data in a d-dimensional space Y which best preserves the manifold's estimated intrinsic geometry. The coordinate vectors y_i for points in the space Y are chosen to minimise the cost function [37].

$$E = \left\| \tau(\mathrm{D_G}) - \tau(\mathrm{D}_Y) \right\|_{L^2} \tag{11}$$

where D_Y represents the matrix of Euclidean distances $\{ d_Y(\mathrm{i,j}) = \left\| y_i - y_j \right\| \}$. $\left\| A \right\|_{L^2}$ denotes the L^2 matrix norm $\sqrt{\sum_{i,j} A_{ij}^2}$. $\tau(\bullet)$ is an operator of converting distances to inner products. The top d eigenvectors of the matrix $\tau(D_G)$ are used to set the coordinates y_i to optimise the above cost function.

The basic feature of ISOMAP is to find a low dimensional embedding of data points. It is a nonlinear and global optimal method since only one free parameter (e.g. ε or K) needs to be optimised. The typical shortcoming of ISOMAP rests on its computational complexity, characterized by full matrix eigenvector decomposition [37].

2.1.4 Minimum Redundancy Maximum Relevance Algorithm

The minimum redundancy maximum relevance (mRMR) algorithm is a well-known and practical dimensionality reduction method of feature selection and fusion. It is achieved by selecting p indexes which have the minimum redundancy with each other and the maximum relevance with different target classes $h = \{h_1, h_2, \ldots, h_K\}$ in m original indexes of N sample data.

The key idea of minimum redundancy is to select the indexes such that they are mutually maximally dissimilar which could be described by mutual information of two variables. Mutual information of two variables x and y involves the joint distribution $p(x, y)$ and their respective marginal probabilities $p(x)$ and $p(y)$ as defined by [38]:

$$I(x,y)=\sum_{i,j} p(x_i,y_j)log\frac{p(x_i,y_j)}{p(x_i)p(y_j)} \tag{12}$$

Suppose S is the selected feature index space, the minimum redundancy condition is given by [38]:

$$minW_I,\ W_I=\frac{1}{|S|^2}\sum_{i,j\in S} I(i,j) \tag{13}$$

where $I(i,j)$ represents $I(g_i,g_j)$ for notational simplicity, and g_i is the feature index. $|S|$ denotes the number of feature index space S.

The idea of maximum relevance is to select the feature indexes that maximise the total relevance of selected feature indexes g_i and target classes $\boldsymbol{h} = \{h_1, h_2, \ldots, h_K\}$. The maximum relevance condition is also formulated by the mutual information as [38]:

$$maxV_I,\ \ V_I=\frac{1}{|S|^2}\sum_{i,j\in S} I(\boldsymbol{h},i) \tag{14}$$

where $I(\boldsymbol{h},i)$ represents $I(h,g_i)$.

Thus, the feature selection of the mRMR algorithm can be achieved by optimising the conditions in Eqs. (13) - (14) simultaneously. These two conditions can be incorporated into a single criterion function (e.g. mutual information difference or mutual information quotient) by [38]:

$$\text{Mutual information difference: } \max(V_I\text{-}W_I) \tag{15}$$

$$\text{Mutual information quotient: } \max(V_I/W_I) \tag{16}$$

As discussed above, these four data fusion techniques are based on different measures (e.g., variance-based, distance-based, and entropy-based, etc.). The advantages and disadvantages of these techniques are compared in Table 2.

Table 2. Performance comparison of different data fusion techniques

Methods	Advantages	Disadvantages
KPCA	Nonlinear transform, computational efficient in a non-iterative way	Performance affected by the selection of kernel functions and reduced feature dimension
LLE	Nonlinear manifold learning, non-iterative way to seek underlying high dimensional structure	Sensitive to noise, ill-conditioned Eigen issue, parameter optimisation required
ISOMAP	Nonlinear manifold learning, global optimisation method in a non-iterative way	High computational cost, overestimated geodesic distance caused by graph discreteness
mRMR	Computationally efficient, reliable estimation of relevance and redundancy of features	Only the feature subset selected, which may not be the optimal representation of data.

2.2 Support Vector Regression Model

Support vector regression (SVR) is based on statistical learning theory for regression analysis [39]. Comparing with other data mining techniques such as artificial neural networks (ANN), it reveals good generalization capability and needs less training samples [40]. SVR transforms the original feature space into a higher dimensional space to determine an optimal hyperplane by maximising the separation distances among the classes. Given an input training data set $\boldsymbol{z} \in \chi$, the transformed higher dimensional feature space can be obtained as:

$$\boldsymbol{z}' = \phi(\boldsymbol{z}) \tag{17}$$

where ϕ is the transformation function. A hyperplane $f(\boldsymbol{z}') = 0$ can be formulated as [40]:

$$f(\boldsymbol{z}') = \boldsymbol{\tau}^T \boldsymbol{z}' + b = \sum_{j=1}^{n} \tau_j z_j' + b = 0 \tag{18}$$

where $\boldsymbol{\tau}$ is a *n*-dimensional vector and *b* is a scalar. The vector $\boldsymbol{\tau}$ and scalar *b* are used to define the position of the separating hyperplane. The hyperplane is built to maximise the distance *D* among the closest classes through the following optimisation.

$$\max_{\boldsymbol{\nu}\in R^n, b\in R} D, \text{ subject to } y_i\left(\boldsymbol{\tau}^T z_i^{'} + b\right) \geq D, \ \forall i \tag{19}$$

where y_i is the class labeler. For example, it is labeled as {-1, 1} for two classes. Taking into account the noise with slack variables ξ_i and error penalty *C*, Eq. (19) can be rewritten as [40]:

$$\begin{aligned} &\min_{\boldsymbol{\nu},\boldsymbol{\xi}\in R^n, b\in R} \left\{ \frac{1}{2}\|\boldsymbol{\tau}\|^2 + C\sum_{i=1}^{N}\xi_i \right\} \\ &\text{subject to } \xi_i \geq 0,\ y_i\left(\boldsymbol{\tau}^T\phi(z_i) + b\right) \geq 1-\xi_i,\ \forall i \end{aligned} \tag{20}$$

The hyperplane can be determined as the following *sign* function (*sgn*(*t*) = 1 for $t \geq 0$, and *sgn*(*t*) = -1 for $t < 0$). The linear decision function is given by:

$$f\left(\boldsymbol{z}\right) = \text{sgn}\left(\sum_{i,j=1}^{N} y_i\alpha_i(\phi^T\left(z_i\right)\phi(z_j)) + b \right) \tag{21}$$

where α is the Lagrange multiplier. The hyperplane function can be determined by kernel function $K(z_i,z_j) = \phi^T(z_i)\phi(z_j)$ by computing the inner products without specifying the explicit form of the transformation function. Different kernels can be formulated such as linear, polynomial, Gaussian RBF, and Sigmoid kernel functions. Accordingly, the associated decision function for regression analysis is expressed as [41]:

$$f\left(\boldsymbol{z}\right) = \text{sgn}\left(\sum_{i,j=1}^{N} y_i\alpha_i K(z_i, z_j) + b \right) \tag{22}$$

The theoretical background of feature fusion techniques and support vector regression discussed here forms the basis of virtual tool wear sensing model formulated in the next section.

3 VIRTUAL TOOL WEAR SENSING FRAMEWORK

It is recognized that indirect sensing techniques (e.g. force, vibration, and acoustic emission, etc.) measure in-process auxiliary parameters during machining operations. The indirect sensing parameters are less accurate to indicate tool conditions, but the rugged senor design makes them more suitable for practical applications. On the other hand, direct sensing techniques (e.g. microscope, CCD camera, etc.) measure actual quantities of tool conditions and have a high degree of accuracy. Due to the practical limitations caused by access problems during machining, illumination and the use of cutting fluid, direct sensing techniques are commonly used for offline measurement or as laboratory techniques.

Utilizing the advantage of indirect sensing, virtual sensing can model the nonlinear dependencies between in-process measurements and actual quantities of tool conditions based on computational models. The accuracy of virtual sensing is expected to be comparable to direct sensing. The rationale of virtual sensing to bridge the gap between indirect sensing and direct sensing is described in Fig. 2.

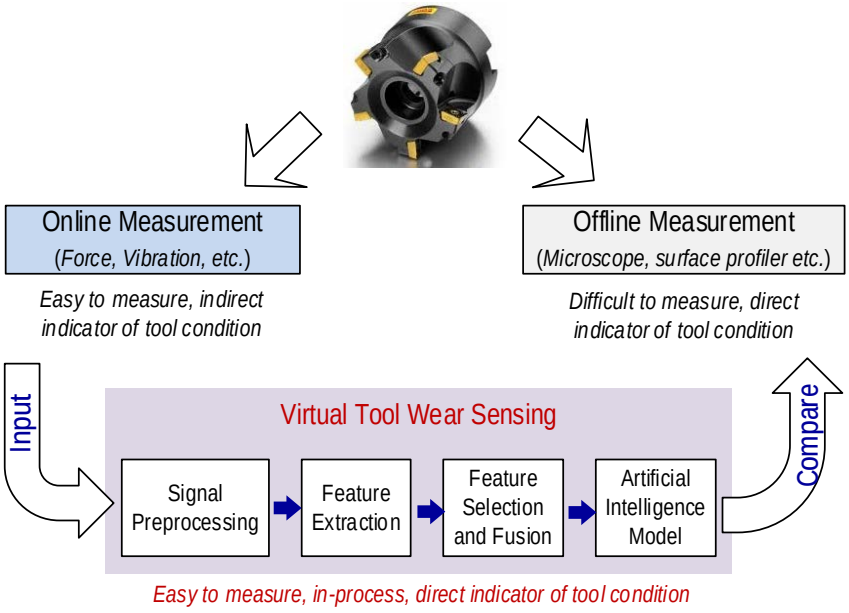

Fig. 2. The rationale of developing virtual tool wear sensing model.

The developed virtual sensing model in this work mainly consists of four modules: (i) a data acquisition system capable of measuring multi-sensory data from machining processes, (ii) a feature extraction module to extract tool health indicators by preprocessing raw noisy measurements, (iii) a feature fusion module to select and fuse the extracted features for dimension reduction, and (iv) a support vector regression based artificial intelligence model to infer tool wear conditions from the fused features, as illustrated in Fig. 3. The developed virtual sensing model is a complement to physical sensing, and can be used for tool wear monitoring and maintenance actions guidance. The deduced virtual sensing parameters are compared to actual quantities of tool wear for performance comparison.

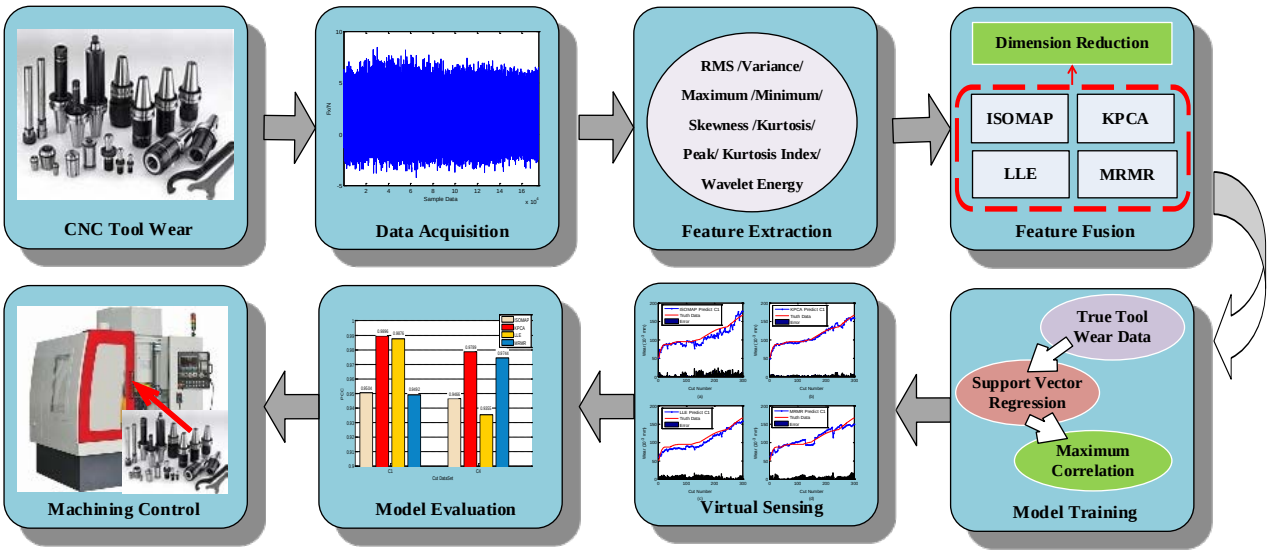

Fig. 3. Diagram of developed virtual tool wear sensing model.

3.1 Multi-Sensory Data Acquisition

During a machining process the action of machining tool removes materials, and meanwhile it causes cutting force variation and machine vibration. Therefore, the sensing parameters including cutting force, vibration, etc. could be used for monitoring machining tool conditions. However, the signals acquired from a manufacturing system in an industrial environment usually

contain high level mechanical, electrical and acoustic noises. On the other hand, strong interference of operating conditions limits the applicability of a single sensor, making it not always fully reflect tool conditions. Thus different modalities of sensors are instrumented.

A multi-sensory data acquisition system is used to identify the dependency between sensing data and tool wear states in machining operations. During the experimental testing, the forces and vibrations in three directions of tangential (main, z-axis), axial (feed, longitudinal, x-axis), and radial (normal, passive, thrust, y-axis) components are measured and stored on a PC for post processing.

3.2 Multi-Domain Feature Extraction

During the tool wear process, online measurement can acquire in-process parameters such as cutting forces and vibrations as indirect indicators for tool wear states. Due to the low signal to noise ratio (SNR) of sensing measurement, it is difficult to model the relationship between raw online measurement and tool wear states. To tackle the problem, effective feature extraction and representation techniques are performed to reduce data dimension without losing tool wear signature.

Different features from statistical, frequency, and time-frequency domains are extracted as shown in Table 3. Six typical statistical features are extracted in statistical domain, including root mean square (RMS), variance, maximum value, skewness, Kurtosis and peak-to-peak. RMS is a measure for the magnitude of a varying quantity. It is also related to the energy of a signal. Kurtosis indicates the spikiness of a signal. Features from the frequency domain provide another perspective of tool wear conditions, and reveal information that are not included in the statistical domain. In frequency domain, spectral skewness and spectral kurtosis are extracted, where $S(f_i)$ is the power spectrum density obtained using the Welch method. In time-frequency domain,

wavelet transform can be used for signal denoising and feature extraction. The wavelet coefficient with higher energy is selected which is related to the characteristic frequency of the machining tool (the number of flutes times the spindle rotating frequency). Thus, the energy of the selected wavelet coefficient is extracted as a feature.

Table 3. List of extracted features.

Domain	Features	Expression
Statistical	RMS	$z_{RMS} = \sqrt{\frac{1}{n}(z_1^2 + z_2^2 + \cdots + z_n^2)}$
	Variance	$z_{\text{var}} = 1/N \sum_N (z_i - \bar{z})^2$
	Maximum	$z_{max} = \max(z)$
	Skewness	$z_{skew} = E\left[\left(\frac{z-\mu}{\sigma}\right)^3\right]$
	Kurtosis	$z_{kurt} = \frac{1}{n}\sum_n \left(\frac{z_i - \mu}{\sigma}\right)^4$
	Peak-to-Peak	$z_{p\text{-}p} = \max(z)\text{-}\min(z)$
Frequency	Spectral skewness	$f_{skew} = \sum_{i=1}^{k} (\frac{f_i - \bar{f}}{\sigma})^3 S(f_i)$
	Spectral Kurtosis	$f_{kurt} = \sum_{i=1}^{k} (\frac{f_i - \bar{f}}{\sigma})^4 S(f_i)$
Time-Frequency	Wavelet energy	$E_{WT} = \sum_{i=1}^{N} wt_{\varphi}^2(i) / N$

3.3 Feature Selection and Fusion

There are an overwhelming number of features extracted from the raw multi-sensory data. In general, multi-sensory features can be viewed as a high-dimensional multivariate matrix composed of several vectors which are formed by different sensory signals. It is not feasible to input the above matrix to a virtual sensing model without dimension reduction because of the curse of dimensionality and the high correlation between vectors. For improved computational

efficiency in virtual tool wear sensing, a proper feature selection and fusion strategy is needed to lower the dimension of a feature space.

By implementing feature selection and fusion algorithm, the complexity of modeling process could be reduced and new feature vectors are reconstructed. Different representative dimensional reduction techniques are investigated for feature selection and fusion, including KPCA, LLE, ISOMAP, and mRMR algorithms. Generally, it is difficult to determine which feature is more sensitive to tool conditions. The goal of feature selection and fusion is to preserve as much of the relevant information as possible by removing redundant or irrelevant information in acquired sensory signals. The top ranked features of these four schemes (e.g. KPCA, LLE, ISOMAP, mRMR, etc.) are then selected and fused into the computational model to infer the actual quantities of tool conditions. Their performance is then evaluated by comparing the estimated tool conditions with actual tool conditions in terms of estimation accuracy.

3.4 SVR based Virtual Sensing Model

Given the complex relationship between fused features and actual quantities of tool conditions, it is difficult to describe it in an explicit analytic form. By exploiting the underlying structure of data measurements, artificial neural network based data driven approach has been investigated to model the nonlinear mapping of signal features of tool wear on artificial intelligence basis. However, virtual sensing based on data driven approaches is limited by several factors: 1) the large amount of training data used to build data-driven models is difficult to collect, since it is costly to obtain labeled data; 2) the lackness of generality or adaptability to extrapolate the conditions beyond the range in which the model was trained leads to model errors. The constraints are amplified by the fact that tools are required to operate under various operating conditions (e.g. varying load and speed).

The recently developed support vector regression model is a novel machine-learning tool which can be used to relieve the above constraints. It is motivated by statistical learning theory and is characterized by the use of non-linear kernels, high generalization ability and the sparseness of the solution with small training sample size. Thus, the support vector regression model is used to investigate the dependency between fused features and tool conditions. The support vector regression model is built and trained, in which two hyper-parameters, the cost parameter C and the Gaussian kernel parameter γ, are selected using the grid search method in the cross validation process to prevent overfitting. The best hyper-parameters are determined in terms of model accuracy through an exhaustive searching over different pairs of these two parameters, which are called grids. The performance of the developed SVR based virtual tool wear sensing model is evaluated based on leave-one-out cross-validation, in which one dataset is chosen for testing while the rest datasets are for model training. Model training and testing are performed repeatedly and different testing dataset is selected each time.

4 EXPERIMENTAL STUDIES

4.1 Experimental Setup

To experimentally analyze the performance of the presented virtual tool wear sensing method, a set of experimental data measured from a high speed CNC machine under dry milling operations was used [42]. A three-flute ball nose tungsten carbide cutter was tested to mill a workpiece (material: stainless steel, HRC52) in a down milling operation. The workpiece had been preprocessed to remove the original skin layer containing hard particles. The speed of the spindle was running at 10,400 rpm while the feed rate was set as 1,555 mm/min in x direction. The depth of cut (radial) was 0.125 mm in y direction while the depth of cut (axial) in z direction was 0.2 mm.

A Kistler quartz 3-component platform dynamometer was mounted between the workpiece and the machining table to measure cutting forces. Three Kistler Piezo accelerometers were mounted on the workpiece to measure the machine tool vibration in x, y, z directions, respectively [42]. During the tool wear test, in-process measurements including force and vibration in three directions (x, y, z) were continuously recorded at the sampling frequency of 50 kHz using DAQ *NI PCI1200*, and were stored in a computer. The flank wear of each individual flute was measured offline using a LEICA MZ12 microscope after finishing each surface which is considered to be one cut number in the following data analysis. The diagram of the experimental setup is shown in Fig. 4.

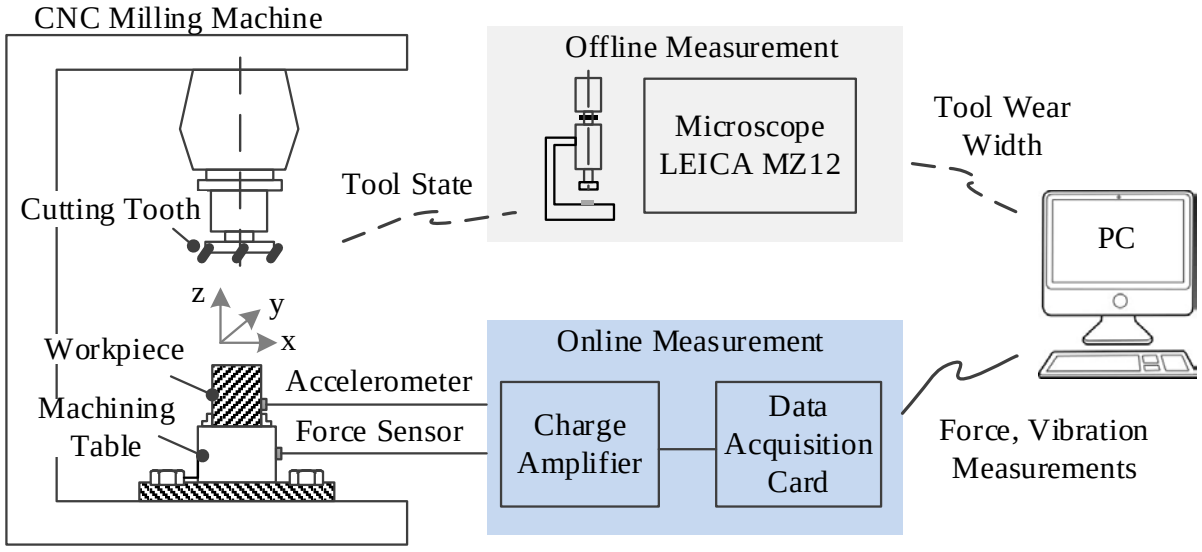

Fig. 4. Schematic diagram of experimental setup.

4.2 Data Processing

A total of around 300 data files (one data file corresponds to one cut number) were collected during the tool life test. The sample measurements of force and vibration under different tool wear conditions are illustrated in Fig. 5. It is obvious that the amplitudes of vibration and force measurements increase as the progressive tool wear conditions. The features discussed in Table 3 are extracted from three dimensional force and vibration measurements and 54 features sets are

obtained. Take the force signal measurements in y direction as an example, the extracted normalized features and actual flank wear measurements are shown in Fig. 6. It is found that some of the extracted features (e.g. RMS, variance, and wavelet energy, etc.) follow the trend of actual tool wear measurements while others (e.g. skewness and spectral kurtosis, etc.) deviate from the trend. The extracted features inevitably present high correlation and redundancy among them. Thus, it is necessary to perform dimension reduction to select dominant features for the subsequent tool wear sensing model.

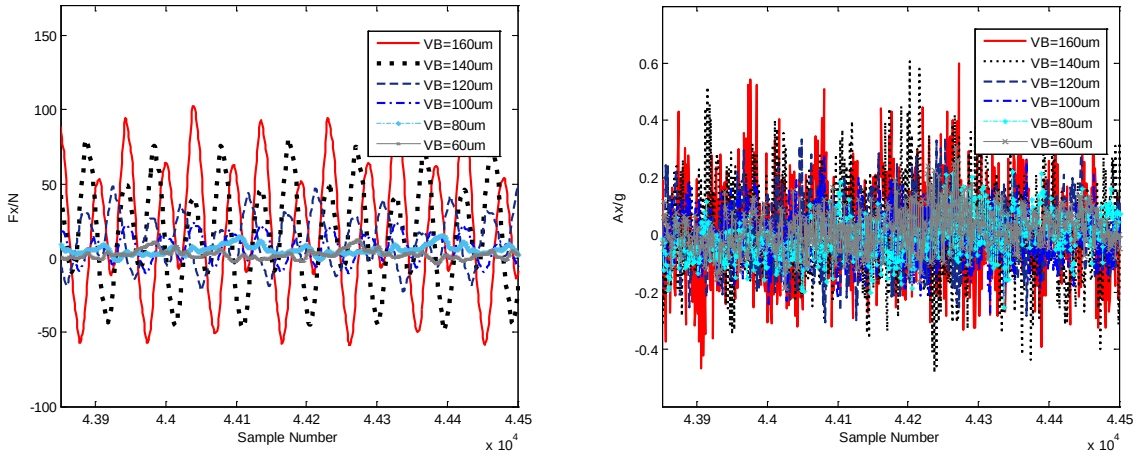

Fig. 5. The sample measurements of force and vibration under different tool conditions.

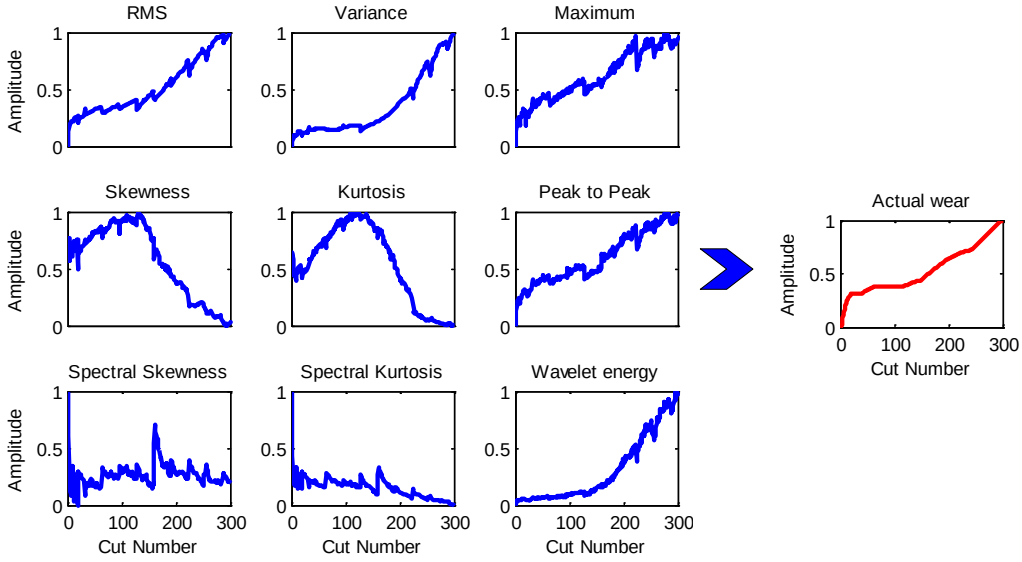

Fig. 6. Extracted features of force signal in y-direction.

The extracted features of multi-sensory measurements are constructed as a matrix. Whitening and eigenvalue decomposition (EVD) are then performed to estimate the number of dominant features, and the results are shown in Fig. 7. It is obvious that the first 11 features with high variances preserve almost 95% of the cumulative variances and thus the 11 features with high variances are selected. Next, the dimensional reduction techniques including KPCA, LLE, ISOMAP, and mRMR are performed to exploit the 11 features by fusing the extracted 54 features.

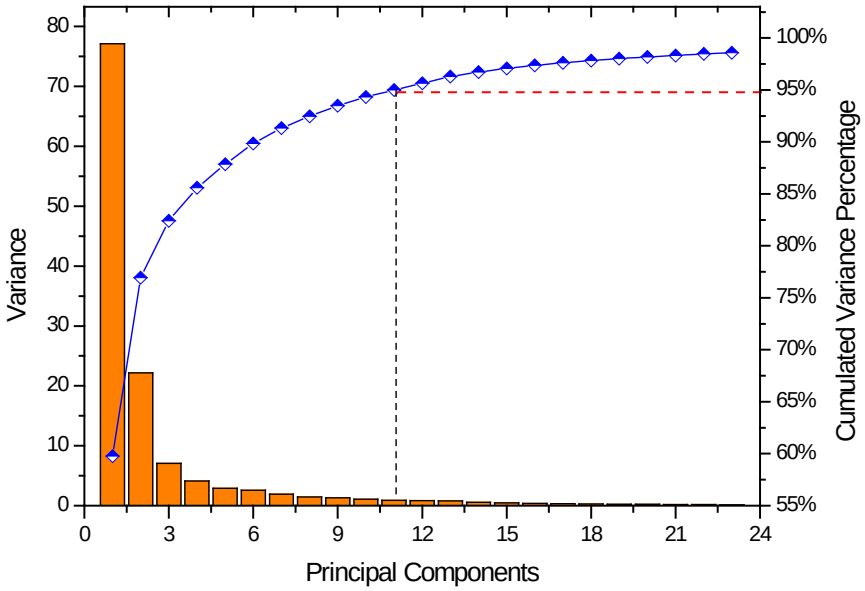

Fig. 7. Dimension reduction through eigenvalue decomposition.

4.3 Performance Evaluation

The selected and fused features obtained by KPCA, LLE, ISOMAP, and mRMR are fed into the support vector regression model to infer the tool wear conditions from the in-process multi-sensory signals. Firstly, the support vector regression model is built and trained in which two parameters, the cost parameter C and the Gaussian kernel parameter γ, are selected using grid search method. A predefined set of these two parameters forms a grid. The best hyper-parameters are determined (C = 2 and γ = 0.11) in terms of model accuracy through an exhaustive searching

in the leave-one-out cross validation process to prevent overfitting. A total of three sets of tool life test data (e.g. *C1*, *C4*, and *C6*, etc.) are available. The predicted tool wear using different virtual sensing schemes including no feature selection, KPCA, LLE, ISOMAP, and mRMR algorithms with the support vector regression model are illustrated in Figs. 8 and 9, respectively. To compare their performance, the actual tool wear condition (flank wear width) measured offline with a microscope is also included. It is found that the predicted tool wear by the virtual sensing model follows the trend of the truth data well.

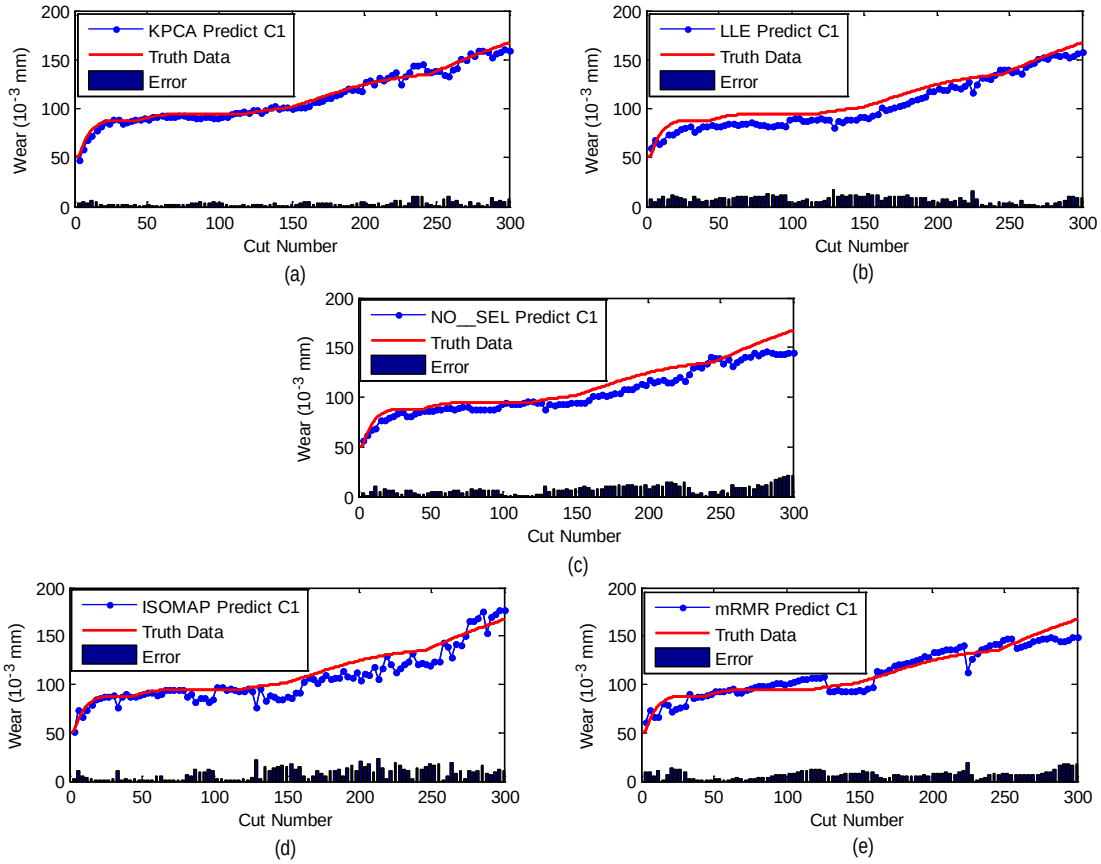

Fig. 8. Performance comparison of different virtual tool wear sensing model using dataset C1.

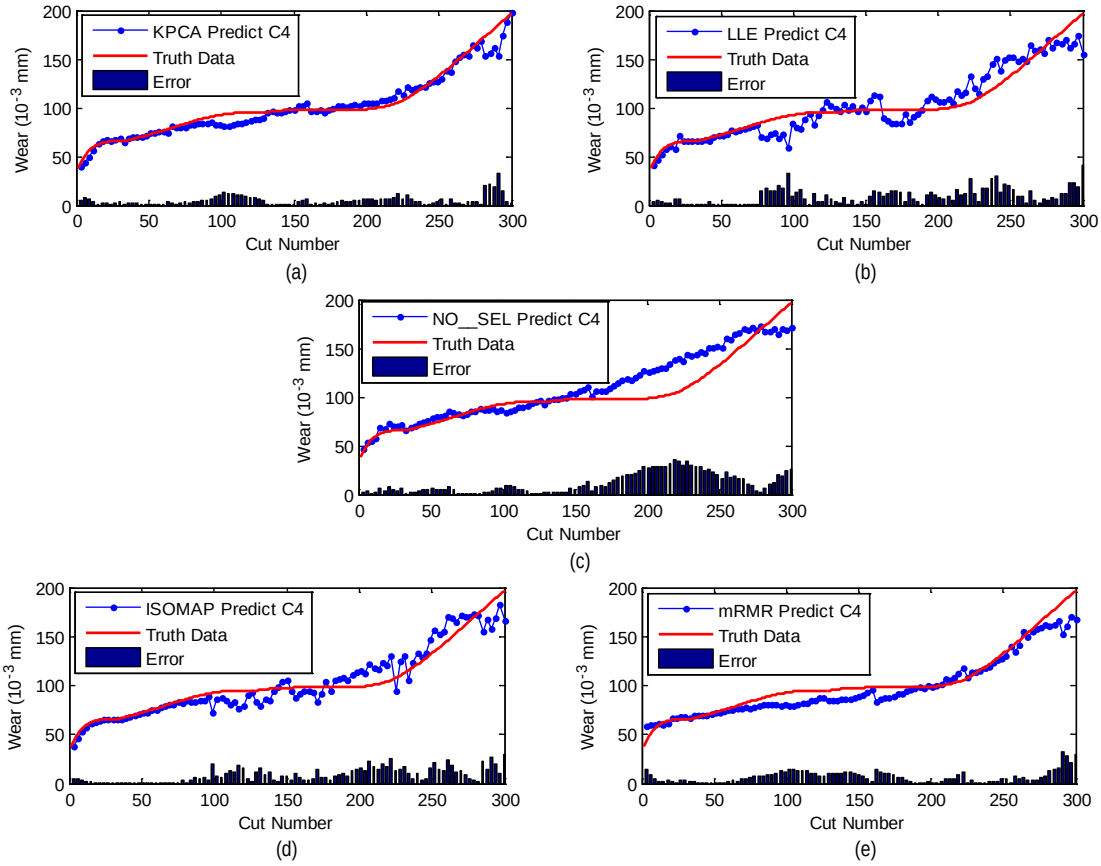

Fig. 9. Performance comparison of different virtual tool wear sensing model using dataset C4.

To quantitatively evaluate the performance of the developed virtual sensing model, different criteria are investigated including Pearson Correlation coefficient (PCC), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). *PCC* is a statistical measure of independence of two or more random variables which is defined as:

$$PCC = \frac{\sum_i (x_i - \overline{x})(\hat{x}_i - \overline{\hat{x}})}{\sqrt{\sum_i (x_i - \overline{x})^2 \sum_i (\hat{x}_i - \overline{\hat{x}})^2}} \tag{23}$$

where x is the actual tool wear width, and $\hat{x}$ is the predicted tool wear width using the virtual sensing model. The model with the highest correlation coefficient is considered as the one of the

best. RMSE is defined as the square root of the average of the square of all difference between estimated tool wear width $\hat{x}$ and actual tool wear width x.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2} \tag{24}$$

MAE is defined as the mean of all absolute difference between estimated tool wear width $\hat{x}$ and actual tool wear width x.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|x_i - \hat{x}_i\right| \tag{25}$$

MAPE is defined as the mean of all absolute percentage differences between estimated tool wear width $\hat{x}$ and actual tool wear width x.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|x_i - \hat{x}_i\right|}{x_i} \tag{26}$$

Next, five different virtual sensing schemes (NO-SEL, KPCA, LLE, ISOMAP, and mRMR) are quantitatively evaluated according to different criteria including PCC, RMSE, MAE, and MAPE. The performance of these five virtual sensing schemes is compared as shown in Fig. 10. The larger the PCC value, the better the model performance, while the less the RMSE/MAE/MAPE value, the better the model performance. In the NO_SEL scheme all features are fed into the support regression model without feature selection, which may cause the overfitting issue, leading to unsatisfactory performance. With the feature fusion techniques, the original features are reduced to alleviate the overfitting issue of the support vector regression model, and the virtual sensing schemes with feature fusion present improved performance. With different test datasets, the performance of the virtual sensing model varies, which may be caused by the varying quality of training datasets and the possible distribution discrepancy between the training and testing domains. The performance of these different virtual sensing schemes is

summarized and compared in Table 4. Overall, the KPCA-SVR virtual sensing model shows the best performance.

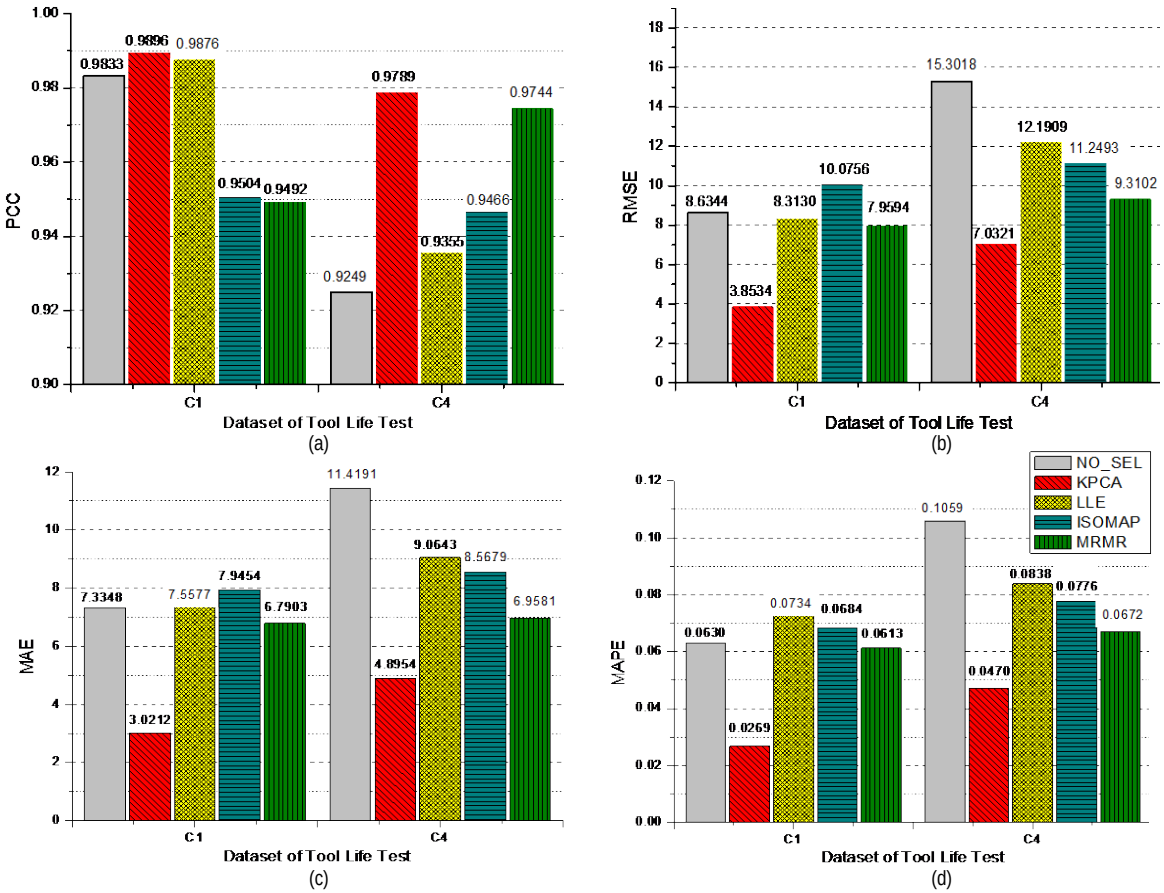

Fig. 10. Performance comparison of five virtual sensing schemes using different criteria.

Table 4. Performance comparison of different virtual sensing schemes

Methods	PCC	RMSE	MAE	MAPE
NO-SEL	0.9541±0.0292	11.9681±3.3337	9.3770±2.0422	0.0845±0.0215
KPCA	0.9843±0.0054	5.4428±1.5894	3.9583±0.9371	0.037±0.01
LLE	0.9616±0.0261	10.2520±1.939	8.311±0.7533	0.0786±0.0052
ISOMAP	0.9485±0.0019	10.6625±0.5869	8.2567±0.3113	0.073±0.0046
mRMR	0.9618±0.0126	8.6348±0.6754	6.8742±0.0839	0.0643±0.003

5 DISCUSSIONS

The developed virtual sensing method takes the extracted features of online measurements as input while the offline measurements are used for performance evaluation. The virtual sensing

method involves data preprocessing, feature extraction, feature fusion, and artificial intelligence model construction. The raw signals acquired in an industrial environment usually present high levels of mechanical, electrical and acoustic noises. On the other hand, the machinery may run under intermittent operating conditions. The signals measured when the cutting tools are in contact with the workpiece carry useful information of cutting conditions. Therefore, appropriate signal preprocessing is needed for signal segmentation to exact signal lobes reflecting material removal. A detailed discussion on signal preprocessing techniques can be found in [43].

The selection of reduced dimensions in feature fusion techniques is another issue. The reduced dimension is selected based on the accumulated contribute rate of feature variances. The threshold is set to be 95% and 11 features are selected since it is believed that the most significant variance can be preserved. Sensitivity analysis on the selection of reduced dimension is further conducted and the results are shown in Fig. 11. It is found that the reduced dimension of 11 features shows the maximum PCC value and the minimum MAE/MAPE value. Although the reduced dimension of 11 features does not perform the best in terms of the RMSE measure, it shows comparable performance. Thus the 11 features are selected by the following feature fusion methods in this paper.

Finally, there exist various artificial intelligence models, such as artificial neural network and support vector regression. Neural network has been extensively used to model the correlation between online sensory signals and tool wear conditions. However, it also suffers from some disadvantages, such as low convergence rate, overfitting issue, poor generalization when few samples are available and the tricky network structure determination. On the other hand, support vector regression is characterized by high generalization ability and the sparseness of the solution with small training sample size. Given the high cost and practical constraints to obtain

data samples in practice, the support vector regression method is used and investigated in this paper. A fair comparison of artificial neural network and support vector regression for virtual sensing model is necessary and will be investigated in future work.

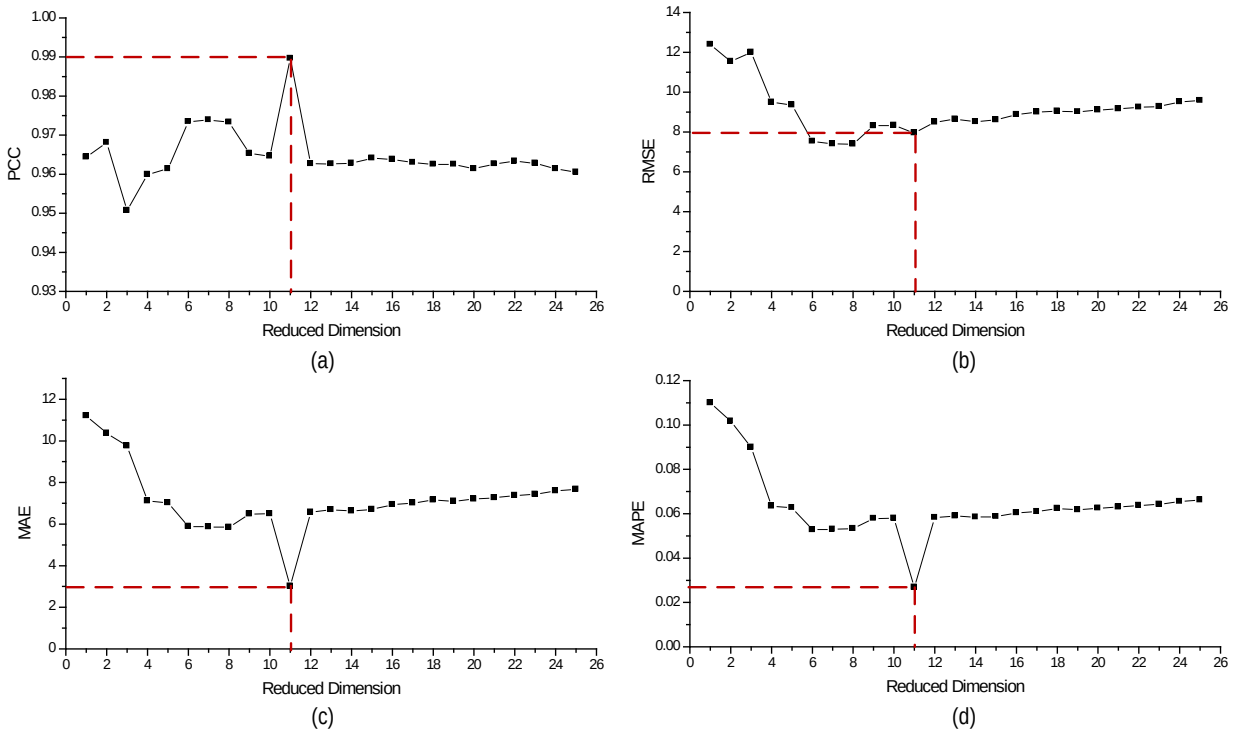

Fig. 11. Sensitive analysis on selection of reduced dimensions for performance evaluation.

6 CONCLUSIONS

To improve the reliability of ubiquitous manufacturing, advanced sensing and signal processing methods are critically needed for effective and efficient fault identification and diagnosis. The developed virtual tool wear sensing framework utilizes in-process multi-sensory measurements to infer the actual tool wear conditions on a support vector regression model basis. Different dimension reduction techniques are investigated for feature selection and fusion to improve the computational efficiency and model accuracy. The following conclusions can be drawn:

1) The multi-sensory fusion based virtual tool wear sensing model bridges the gap between indirect sensing and direct sensing for tool condition monitoring and prediction.
2) A variety of dimension reduction techniques including KPCA, LLE, ISOMAP, and mRMR algorithms have been investigated for multisensory feature and fusion, and experimental results show that KPCA performs best.
3) The effectiveness of the developed virtual tool wear sensing model is validated using a set of experimental machining tool life tests, and its performance is comparable to the costly offline instrumentation (e.g. microscope).

For future work, a variety of experimental tests under different operating conditions will be performed to evaluate the robustness of the presented virtual sensing model. The developed virtual sensing model will be incorporated in a CNC machine control system to improve control accuracy and machine reliability.

ACKNOWLEDGMENTS

This research acknowledges the financial support provided by National Science foundation of China (No. 51504274), Science Foundation of China University of Petroleum, Beijing (No. 2462014YJRC039 and 2462015YQ0403). Valuable comments from anonymous reviewers are greatly appreciated to improve the paper's quality.

REFERENCES

[1] M. Friedewald, O. Raabe, Ubiquitous computing: an overview of technology impacts, Telematics and Informatics, 28(2011), pp. 55-65.
[2] R. Dubey, A. Gunasekaran, A. Chakrabarty, Ubiquitous manufacturing: overview, framework, and further research directions, International Journal of Computer Integrated Manufacturing, 2015, DOI: 10.1080/0951192X.2014.1003411.

[3] Y. Zhang, G.Q. Huang, T. Qu, O. Ho, S. Sun, Agent-based smart objects management system for real-time ubiquitous manufacturing, Robotics and Computer-Integrated Manufacturing, 27(2011), pp. 538-549.
[4] Y. Zhang, G.Q. Huang, S. Sun, T. Yang, Multi-agent based real-time production scheduling method for radio frequency identification enabled ubiquitous shopfloor environment, Computers & Industrial Engineering, 76(2014), pp. 89-97.
[5] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing, 20(2016), pp. 1483-1510.
[6] S. Kurada, C. Bradley, A review of machine vision sensors for tool condition monitoring, Computers in Industry, 34(1997), pp. 55-72.
[7] R. Teti, K. Jemielniak, G. O'Donnell, D. Dornfeld, Advanced monitoring of machining operations, CIRP Annals – Manufacturing Technology, 59(2010), pp. 717-739.
[8] R. Mejia-Gutierrez, G. Osorio-Gomez, D. Rios-Zapata, D. Zuluaga-Holguin, Ubiquitous conceptual design of a ubiquitous application: a textile SME case study for real time manufacturing monitoring, Computer-Aided Design, 59(2015), pp. 214-228.
[9] G. Sagl, T. Blaschke, E. Beinat, and B. Resch, Ubiquitous Geo-sensing for context-aware analysis: exploring relationships between environmental and human dynamics, Sensors, 12(2012), pp. 9800-9822.
[10] J. Lloret, E. Macias, A. Suarez, and R. Lacuesta, Ubiquitous monitoring of electrical household appliances, Sensors, 12(2012), pp. 15159-15191.
[11] Y. Zhang, L. Sun, H. Song, X. Cao, Ubiquitous WSN for healthcare: recent advances and future prospects, IEEE Internet of Things Journal, 1(4)(2014), pp. 311-318.
[12] S. Kurada, C. Bradley, A machine vision system for tool wear assessment, Tribology International, 30(4)(1997), pp. 295-304.
[13] C. Zhang, J. Zhang, On-line tool wear measurement for ball-end milling cutter based on machine vision, Computers in Industry, 64(2013), pp. 708-719.
[14] S. Dutta, S.K. Pal., S. Mukhopadhyay, R. Sen, Application of digital image processing in tool condition monitoring: a review, CIRP Journal of Manufacturing Science and Technology, 6(2013), pp. 212-232.
[15] W.H. Wang, G.S. Hong, Y.S. Wong, and K.P. Zhu, Sensor fusion for online tool condition monitoring in milling, International Journal of Production Research,45(21)(2007), pp. 5095-5116.
[16] M. Milfelner, F. Cus, J. Balic, An overview of data acquisition system for cutting force measuring and optimization in milling, Journal of Materials Processing Technology, 164-165(2005), pp. 1281-1288.
[17] B. Chen, X. Chen, B. Li, Z. He, H. Cao, G. Cai, Reliability estimation for cutting tools based on logistic regression model using vibration signals, Mechanical Systems and Signal Processing, 25(2011), pp. 2526-2537.
[18] J. Zhou, C.K. Pang, Z. Zhong, F.L. Lewis, Tool wear monitoring using acoustic emissions by dominant-feature identification, IEEE Transactions on Instrumentation and Measurement, 60(2)(2011), pp. 547-559.
[19] B. H. Freyer, P.S. Heyns, N.J. Theron, Comparing orthogonal force and unidirectional strain component processing for tool condition monitoring, Journal of Intelligent Manufacturing, 25(2014), pp. 473-487.

[20] A.A.D. Sarhan, A. Matsubara, Investigation about the characterization of machine tool spindle stiffness for intelligent CNC end milling, Robotics and Computer-Integrated Manufacturing, 34(2015), pp. 133-139.
[21] D.R. Salgado, F.J. Alonso, An approach based on current and sound signals for in-process tool wear monitoring, International Journal of Machine Tools & Manufacture, 47(2007), pp. 2140-2152.
[22] M.T. Tham, G.A. Montague, J.A. Morris, P.A. Lant, Soft sensors for process estimation and inferential control, Journal of Process Control, 1(1)(1991), pp. 3-14.
[23] C.D. Petersen, R. Fraanje, B.S. Cazzolato, A.C. Zander, C.H. Hansen, A Kalman filter approach to virtual sensing for active noise control, Mechanical Systems and Signal Processing, 22(2008), pp. 490-508.
[24] J.J.W. Cheng, T.C. Chao, L.H. Chang, B.F. Huang, A model-based virtual sensing approach for the injection molding process, Polymer Engineering and Science, 44(9)(2004), pp. 1605-1614.
[25] J. Ploennigs, A. Ahmed, B. Hensel, P. Stack, K. Menzel, Virtual sensors for estimation of energy consumption and thermal comfort in buildings with underfloor heating, Advanced Engineering Informatics, 25(4)(2011), pp. 688-698.
[26] M. Ragaglia, A.M. Zanchettin, L. Bascetta, P. Rocco, Accurate sensorless lead-through programming for lightweight robots in structured environments, Robotics and Computer-Integrated Manufacturing, 39(2016), pp. 9-21.
[27] H.C. Huang, Y.C. Lin, M.H. Hung, C.C. Tu, F.T. Cheng, Development of cloud-based automatic virtual metrology system for semiconductor industry, Robotics and Computer-Integrated Manufacturing, 34(2015), pp. 30-43.
[28] A. Bustillo, M. Correa, and A. Renones, A virtual sensor for online fault detection of multitooth-tools, Sensors, 11(2011), pp. 2773-2795.
[29] C.J. Li, T.C. Tzeng, Multimilling-insert wear assessment using non-linear virtual sensor, time-frequency distribution and neural networks, Mechanical Systems and Signal Processing, 14(6)(2000), pp. 945-957.
[30] T. Chen, Y.C. Wang, Estimating simulation workload in cloud manufacturing using a classifying artificial neural network ensemble approach, Robotics and Computer-Integrated Manufacturing, 38(2016), pp. 42-51.
[31] D. Shi, N.N. Gindy, Tool wear predictive model based on least squares support vector machines, Mechanical Systems and Signal Processing, 21(2007)1799-1814.
[32] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak, Health assessment and life prediction of cutting tools based on support vector regression, Journal of Intelligent Manufacturing, 26(2015)213-223.
[33] D. Brezak, D. Majetic, T. Udiljak, J. Kasac, Tool wear estimation using an analytic fuzzy classifier and support vector machines, Journal of Intelligent Manufacturing, 23(2012)797-809.
[34] Q. He, F. Kong, R. Yan, Subspace-based gearbox condition monitoring by kernel principal component analysis, Mechanical Systems and Signal Processing, 21(2007), pp. 1755-1772.
[35] B. Scholkopf, A. Smola, K.R. Muller, Nonlinear component analysis as a kernel eigenvalue, Neural Computation, 10(1998), pp. 1299-1319.
[36] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science, 290(2000), pp. 2323-2326.

[37] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science, 209(2000), pp. 2319-2323.
[38] H. Peng, F. Long, and C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8)(2005), pp. 1226-1238.
[39] V.N.Vapnik, The nature of statistical learning theory. New York: Springer, 1999.
[40] A. Widodo, B.S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mechanical Systems and Signal Processing, 21(2007), pp. 2560–2574
[41] J. Wang, S. Liu, R.X. Gao, R. Yan, Current envelope analysis for defect identification and diagnosis in induction motors, Journal of Manufacturing Systems, 31(4)(2012), pp. 380-387.
[42] X. Li, B.S. Lim, J.H. Zhou, S. Huang, S.J. Phua, K.C. Shaw, M.J. Er, Fuzzy neural network modelling for tool wear estimation in dry milling operation. Annual Conference of the Prognostics and Health Management Society, San Diego, CA, September 27-October 1, 2009; pp. 1-11.
[43] N. Ghosh, Y.B. Ravi, A. Patra, S. Mukhopadhyay, S. Paul, A.R. Mohanty, A.B. Chattopadhyay, Estimation of tool wear during CNC milling using neural network-based sensor fusion, Mechanical Systems and Signal Processing, 21(2007), pp. 466-479.