Monitoring of Signals from Manufacturing Processes Using the Karhunen-Loève Transform [1]

I. Y. Tumer
Research Scientist
Computational Sciences
NASA Ames Research Center
Moffett Field, CA 94035-1000
itumer@ptolemy.arc.nasa.gov

K. L. Wood
Associate Professor
Mechanical Engineering
The University of Texas
Austin, TX, 78712-1063
wood@mail.utexas.edu

I. J. Busch-Vishniac
Dean
Whiting School of Engineering
Johns Hopkins University
Baltimore, MD, 21218
ilenebv@jhu.edu

Abstract

Monitoring the condition of signals during manufacturing can provide valuable information about the status of the parts being produced, as well as that of the components of the manufacturing machine. To this end, we introduce a monitoring technique based on the Karhunen-Loève transform, which decomposes measured manufacturing signals into decorrelated components in the form of fundamental eigenfunctions. The isolated eigenfunctions are monitored by means of the corresponding coefficient vectors, which indicate stationary and nonstationary changes in the deterministic and stochastic fundamental patterns. In this paper, the mathematical foundations of the technique are explored to form a good understanding of the mechanics of the transformation. This understanding is achieved with simple illustrative examples in the discrete and continuous domains, followed by an example of multicomponent signal decomposition by means of numerical simulations. The monitoring potential of the technique is illustrated with an example from a milling process, where the main modes are extracted by means of the eigenvectors and coefficient vectors.

1 Introduction

Signals from manufacturing processes (e.g., vibration from machine components and surface profiles from manufactured parts) provide a good indication of the machine and manufacturing condition [11, 18, 27, 42]. Monitoring of deviations from ideal conditions during part manufacturing can result in considerable cost savings, as well as avoid premature failures while the end part is in operation. This work currently focuses on monitoring the condition of part surfaces. To achieve monitoring, sample surface profile measurements are assumed to be collected on a regular basis during part production. Any faults or deviations in the dynamics of the machine or its submechanisms are assumed to leave a "fingerprint" on the surface being manufactured [39]. The collected surface profiles are then analyzed using signal processing tools to extract information about the fault condition of the part.

[1] In Mechanical Systems and Signal Processing Journal, 2000, 14(6), pp.1011-1026.

The field of condition monitoring constantly searches for better tools to accomplish the above tasks accurately. The most commonly used tools often result in obscured information about the fault condition and result in frequent false alarms, delayed warnings about the fault condition, or insensitivity to a legitimate fault condition [32, 39]. Robustness of such tasks will enable the automation of part production and decision making about the manufacturing condition, hence assuring the fault-free production of high-quality parts [12, 31, 43]. In this work, we introduce the use of a promising mathematical transform to decompose signals into individual components which make physical sense, and monitor each component separately for faults.

1.1 Characteristics of Manufacturing Signals

We view manufacturing as a system whose outputs are signals to be monitored regularly for faults. Signals can be categorized as deterministic or stochastic. Deterministic signals may be detected easily, since they follow a known model by which exact future values can be predicted. Stochastic signals, on the other hand, do not typically follow a known model, and their structure and future values can only be described by probabilistic statements [6, 9]. Stochastic signals can exhibit stationary and/or nonstationary characteristics. The probabilistic laws describing stationary random signals are time-invariant, i.e., the statistical properties of the signal do not change with time. Nonstationary signals, where statistical parameters change significantly with time, can typically be regarded as deterministic factors operating on otherwise stationary random processes [6, 8].

In this work, the focus is on signal types that are typically encountered in manufacturing processes. Most manufacturing processes generate periodic waveforms that are indicative of many potential error sources [26, 39]. Examples are the tool marks from a turning process, feed marks from a milling process, or roller chatter marks from a Selective Laser Sintering process [36, 37]. Stationary or nonstationary changes in these periodic waveforms can be indicative of potential or already existing faults in the machine or material. The appearance of additional periodic components (e.g., harmonics) can be indicative of inherent errors in the manufacturing machine. Component surfaces may contain linear trends such as slopes and offset changes due to impulsive forces during machining (e.g., surface hardness variations, chip breakage, and tool wear).

1.2 Transforms for Manufacturing Signals

The characterization of manufacturing outputs in terms of "signals" necessitates the use of advanced mathematical transforms to obtain a decorrelated decomposition [1, 39]. In this paper, we focus on such mathematical transforms, and aim to find the best candidate for our manufacturing application. The use of advanced mathematical transforms for fault detection has been the focus of many research efforts. The Fourier transform is such a transform, resulting in a decomposition into different frequency components based on sines and cosines as basis vectors [7, 19, 39]. However, the decomposition is not accurate in the

presence of nonstationary features, including linear trends and offset changes. When dealing with real data, nonstationarities and linear trends appear falsely as additional low-frequency components, and modify the magnitudes of the remaining frequency components. The most popular alternative of the Fourier transform is a windowed version of the power spectrum, namely, the short-time Fourier transform [39, 41]. This technique gives a picture of the frequency decomposition of the signal in different time windows, and hence presumably can analyze time-varying signals. However, stationarity still has to be assumed within the short time window, which causes a problem with the frequency/time resolution tradeoff [40].

Popular alternatives to the Fourier transform are the Wigner-Ville distribution, the wavelet transform, and the higher-order spectral transform. The Wigner-Ville transformation typically introduces redundant features when dealing with multicomponent signals [3, 10, 28, 33, 40]. This redundancy tends to obscure the significant fault features. Wavelet transforms provide an exciting tool which eliminates the resolution and accuracy problems of other time-frequency decompositions, while providing a more flexible time-scale resolution [15, 17, 20, 22, 23]. However, in the standard formulation of wavelet transforms, it is difficult to attribute physical meaning to the features (wavelet coefficients) extracted from the wavelet decomposition. Higher-order spectral transforms are used to detect unusual frequency components that would not have been detected with the second-order Fourier spectrum, maintaining phase information in addition to the magnitude information provided by the second-order spectral methods [5, 13, 24, 29]. However, it is often difficult to isolate and interpret the features corresponding to nonstationary signals. In addition to these transforms, there exist other transforms, including Gabor transforms, Walsh transforms, and Haar transforms [21, 39], but these are simply extensions of the Fourier, Wigner-Ville, and wavelet transforms, so their utility is similar.

1.3 Proposed Approach

In this work, we investigate the potential of an alternative transform, namely, the Karhunen-Loève (KL) transform [35, 37, 38]. This orthogonal transform decomposes the signals into completely decorrelated components in the form of empirical basis functions that contain the majority of the variations in the original data. The following sections demonstrate the mechanics of the KL transformation, investigating the decomposition properties of the transform using simple analytical examples, followed by a numerical investigation of the properties. The applicability of the method is demonstrated with a brief example from a milling process. The purpose is to demonstrate the mechanics of the transformation and show feasibility for a KL-based monitoring technique. If the decomposition results in accurate components which make physical sense, then the KL transform can be extended to condition monitoring of manufacturing signals.

2 KL-Based Condition Monitoring

Most condition monitoring techniques depend on the extraction of previously known patterns in signals, such as linear trends and sinusoidal components, by projecting the data onto pre-determined components; such is the case with Fourier-based methods, for example, where the idea is to expand the data in terms of sines and cosines, following the removal of linear effects by regression.

What if we approached this problem from a different point of view? Instead of identifying patterns whose characteristics we know, what if we could identify any type of pattern with a single method, no matter what the pattern is? This would require the use of a transform where the data is not projected onto known functions, such as sines and cosines. With such a tool, not only could we decompose surface signals into their significant components, but we could also use this tool to monitor changes in each of these components with the purpose of diagnosing fault mechanisms in the manufacturing system. Such a transform would have the satisfy the main requirements for implementation in manufacturing, which are: (1) ease of interpretation; (2) clarity of the information from the transformation. These requirements imply the necessity for an unambiguous picture of the signal's content and its changes over time.

With these thoughts in mind, we have explored the existence of a mathematical transform that would provide us with these "idealistic" results. The Karhunen-Loève (KL) transform is such a candidate. The transform projects the data onto empirically-determined functions which contain the maximum variability in the data, rather than pre-determined functions. In the following, we investigate the mechanics of the transform with a systematic study of its decomposition properties, with the purpose of gaining a better understanding of the outputs and the procedure. This understanding will assure us that the requirements for implementing such a monitoring scheme based on the KL transform are satisfied.

2.1 Karhunen Loève Transform: Literature Background

The KL transform is used in a variety of signal processing applications, mainly as a means of detecting "dominant" characteristics in a set of signals [2, 4, 16, 25, 30]. Two example applications are in pattern recognition of faces, and understanding of turbulent flow. Sirovich et al. apply the Karhunen-Loève transform to pattern recognition, attacking the general problem of characterizing, identifying, and distinguishing individual patterns drawn from a well-defined class of patterns [30]. The patterns are pictures of faces of people, taken from a random sample of males. They show that a compressed version of the original faces can be reconstructed with reduced dimensionality to generate recognizable faces. Ball et al. use the Karhunen-Loève transform in the analysis of low-Reynolds number turbulent channel flow. Snapshots of the velocity field $u(x,t)$ are taken and the entire ensemble of velocity measurements are decomposed into spatial modes (basis eigenvectors) and time-dependent amplitude coefficients [4]. They show that "coherent" modes can be detected in turbulence and studied to understand the mechanics of turbulence. In manufacturing, the

applications of KL are rare. One example is the use of Principal Components Analysis (PCA, used in the statistical field, equivalent to the KL transform) as a means of multivariate statistical process control [25]. In this work, a multitude of parameters (such as color, texture, strength, temperature, speed, etc.) is measured in a plant and considered simultaneously to detect variations.

2.2 Karhunen-Loève Transform: Mathematical Background

To enable the use of the KL transform for effective condition monitoring, we claim that the KL transform results in individual patterns, regardless of the type of signal, as long as the variability information is significant. These decomposed patterns correspond to the fundamental modes in the manufacturing system and can be monitored individually for potential faults. Given the individual fundamental patterns of the measured signal, we can then use the corresponding coefficient vectors to monitor the change in time of each pattern. This result will enable the monitoring of fault occurrences in a manufacturing process [37, 38].

The KL transform decomposes multicomponent signals into basis functions which represent the highest variability in the signals. Let us assume that a random process $x(t)$, defined in the time domain (or spatial domain) $(0, T)$, can be expressed as a linear combination of orthogonal functions [14]:

$$\mathbf{x}(t) = \sum_{i=1}^{n} \mathbf{y}_i \varphi_i(t), \tag{1}$$

where the orthogonal functions $\varphi_i(t)$ are deterministic functions, and the ith coefficient $\mathbf{y}_i$ is a random variable. Equation 1 is the KL expansion of the original data, composed of a linear combination of the basis functions and the feature coefficients in the transformed domain.

When applying the KL decomposition to real data, the input functions become discretized versions of the continuous function $x(t)$. The basis functions are computed as the eigenvectors of the discretized version of the covariance function of the inputs, defined as $R(t, \tau) = E[\mathbf{x}(t)\mathbf{x}^*(\tau)]$. The coefficients are computed as the projection of the input functions onto the basis functions using $\int_{i=0}^{T} x(t)\phi(t)dt$, and indicate the weight of each input function in the new transform domain, spanned by the basis functions [14].

To investigate the ability of the KL transform to decompose signals, we assume that a general function $g(x, t)$ represents a manufacturing signal, such as a surface profile. This function is composed of a multitude of functions $f_1(x, t)$, $f_2(x, t)$, etc. In reality, the exact shape of the functions f_i, and the exact nature of the interactions between these functions are not known *a priori*. The decomposition of $g(x, t)$ into individual functions f_i will identify the individual dominant modes.

In the following, we use the well-known mathematics of the KL transformation to illustrate the mechanics of signal decomposition using a set of two-dimensional input vectors and illustrate how the outputs of the KL transformation are computed and what they signify. A good understanding of the mechanics of the transformation is crucial to the successful implementation of the method in a manufacturing environment.

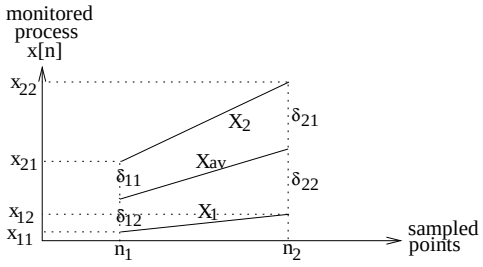


Figure 1: Simple Case: Discretized Linear Patterns

This example serves as a precursor to the general case of N-dimensional and multicomponent functions, which is illustrated next in the continuous domain, using a signal composed of a linear function and two sinusoidal functions. A numerical example follows to illustrate the decomposition using a high-dimensional input matrix, for a set of multicomponent signals composed of a linear and two sinusoidal patterns.

2.2.1 Discrete-Domain Decomposition

Two sample inputs $\mathbf{X_1}$ and $\mathbf{X_2}$, each with two sample points $(x_{11}\, x_{12})^T$ and $(x_{21}\, x_{22})^T$, respectively, are collected, as shown in Figure 1 (M=2, N=2, where M is the number of input vectors or snapshots, and N is the number of sample points in each vector). The inputs represent discretized versions of linear patterns with increasing slopes. The mean input vector is $\mathbf{X}_a = (x_{a1}\, x_{a2})^T$. δ_{ij} is assumed to be the deviation of the ith sampled point of the jth input from the mean. In this case, since we only have two inputs, the average input is equidistant from each input, i.e., $\delta_{11} = \delta_{12}$ and $\delta_{21} = \delta_{22}$, as shown in Figure 1.

Eigenvector and Eigenvalue Computation Let $\mathbf{X_1} = [x_{11}\, x_{12}]^T$ and $\mathbf{X_2} = [x_{21}\, x_{22}]^T$ be two discretized patterns with two points each, representing the input vectors. The mean input vector is $\mathbf{X}_a = (x_{a1}\, x_{a2})^T = [\frac{x_{11}+x_{21}}{2}\ \ \frac{x_{12}+x_{22}}{2}]^T$. Subtracting the mean vector from the input vectors, we obtain the zero-mean input vectors, used to compute the covariance matrix: $\mathbf{Z_1} = \mathbf{X_1} - \mathbf{X}_a = [x_{11} - x_{a1}\ x_{12} - x_{a2}]^T$, and, $\mathbf{Z_2} = \mathbf{X_2} - \mathbf{X}_a = [x_{21} - x_{a1}\ x_{22} - x_{a2}]^T$. Assuming δ_{ij} deviations of the input vectors from the mean vector, the zero-mean input vectors become: $\mathbf{Z_1} = [-\delta_{11}\ \ -\delta_{21}]^T$, and, $\mathbf{Z_2} = [\delta_{11}\ \delta_{21}]^T$. The data matrix becomes $\mathbf{Z} = [\mathbf{Z}_1\, \mathbf{Z}_2]$. The covariance matrix is computed using:

$$\hat{\mathbf{S}} = \frac{1}{2}\sum_{i=1}^{2}\mathbf{Z}_i\mathbf{Z}_i^T = \frac{1}{2}\left[\begin{bmatrix} -\delta_{11} \\ -\delta_{21} \end{bmatrix}\begin{bmatrix} -\delta_{11} & -\delta_{21} \end{bmatrix} + \begin{bmatrix} \delta_{11} \\ \delta_{21} \end{bmatrix}\begin{bmatrix} \delta_{11} & \delta_{21} \end{bmatrix}\right] = \begin{bmatrix} \delta_{11}^2 & \delta_{11}\delta_{21} \\ \delta_{21}\delta_{11} & \delta_{21}^2 \end{bmatrix} \tag{2}$$

The eigenvalues are computed using $det(\mathbf{S} - \lambda\mathbf{I}) = 0$, which becomes:

$$\begin{vmatrix} \delta_{11}^2 - \lambda & \delta_{11}\delta_{21} \\ \delta_{21}\delta_{11} & \delta_{21}^2 - \lambda \end{vmatrix} = 0 \Leftrightarrow (\delta_{11}^2 - \lambda)(\delta_{21}^2 - \lambda) - \delta_{11}^2\delta_{21}^2 = 0 \Leftrightarrow \lambda^2 - (\delta_{11}^2 + \delta_{21}^2)\lambda = 0 \tag{3}$$

The solution of this characteristic polynomial in λ is $\lambda = 0$ or $\lambda = \delta_{11}^2 + \delta_{21}^2$, which are the eigenvalues. The corresponding eigenvector is computed using $(\mathbf{S} - \lambda\mathbf{I})\mathbf{\Phi}_i = 0$, which becomes:

$$(\mathbf{S} - \lambda\mathbf{I})\mathbf{\Phi}_i = 0 = \begin{bmatrix} \delta_{11}^2 - (\delta_{11}^2 + \delta_{21}^2) & \delta_{11}\delta_{21} \\ \delta_{21}\delta_{11} & \delta_{21}^2 - (\delta_{11}^2 + \delta_{21}^2) \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} -\delta_{21}^2 & \delta_{11}\delta_{21} \\ \delta_{21}\delta_{11} & -\delta_{11}^2 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} = -\delta_{21}^2\varphi_1 + \delta_{11}\delta_{21}\varphi_2 = 0 \Leftrightarrow \varphi_1 = \frac{\delta_{11}}{\delta_{21}}\varphi_2$$

Since the eigenvector has to be orthonormal, the orthonormality condition $\mathbf{\Phi}^T\mathbf{\Phi} = \mathbf{I}$ implies $\varphi_1^2 + \varphi_2^2 = 1$, resulting in $\varphi_2 = \frac{\delta_{21}}{\sqrt{\delta_{11}^2+\delta_{21}^2}}$ and $\varphi_1 = \frac{\delta_{11}}{\sqrt{\delta_{11}^2+\delta_{21}^2}}$.

The resulting eigenvector is dependent only on the deviations of each input vector from the mean vector. The eigenvector from above is $\mathbf{\Phi} = [\varphi_1\, \varphi_2]^T = [\frac{\delta_{11}}{\sqrt{\delta_{11}^2+\delta_{21}^2}}\ \frac{\delta_{21}}{\sqrt{\delta_{11}^2+\delta_{21}^2}}]^T$, which has coordinate values that are only dependent on the deviations δ_{11} and δ_{21} of the input vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively, from the mean vector $\mathbf{X}_a$.

Coefficient Vector Computation The coefficient vector $\mathbf{Y} = [y_1\, y_2]^T$, corresponding to the fundamental eigenvector $\mathbf{\Phi} = [\varphi_1\, \varphi_2]^T$, indicates the change in slope between the two input vectors $\mathbf{X_1}$ and $\mathbf{X_2}$. The coefficient vector $\mathbf{Y} = [y_1\, y_2]^T$ corresponding to the eigenvector $\mathbf{\Phi} = [\varphi_1\, \varphi_2]^T = [\frac{\delta_{11}}{\sqrt{\delta_{11}^2+\delta_{21}^2}}\ \frac{\delta_{21}}{\sqrt{\delta_{11}^2+\delta_{21}^2}}]^T$ is computed using $\mathbf{Y} = \mathbf{\Phi}^T\mathbf{Z}$, which becomes:

$$\begin{bmatrix} y_1 & y_2 \end{bmatrix} = \begin{bmatrix} \frac{\delta_{11}}{\sqrt{\delta_{11}^2+\delta_{21}^2}} & \frac{\delta_{21}}{\sqrt{\delta_{11}^2+\delta_{21}^2}} \end{bmatrix} \begin{bmatrix} -\delta_{11} & \delta_{11} \\ -\delta_{21} & \delta_{21} \end{bmatrix} = \begin{bmatrix} -\sqrt{\delta_{11}^2 + \delta_{21}^2}\sqrt{\delta_{11}^2 + \delta_{21}^2} \end{bmatrix} \tag{5}$$

The components y_1 and y_2 can be plotted with respect to the number of input vectors to indicate a positive change between the two input vectors. The change between the two input vectors is equal to the difference $\mathbf{Z_2} - \mathbf{Z_1}$, which is equal to the sum of the δ_{ij} deviations from the mean vector:

$$\mathbf{Z_2} - \mathbf{Z_1} = \begin{bmatrix} \delta_{11} \\ \delta_{21} \end{bmatrix} - \begin{bmatrix} -\delta_{11} \\ -\delta_{21} \end{bmatrix} = \begin{bmatrix} 2\delta_{11} \\ 2\delta_{21} \end{bmatrix} \tag{6}$$

Since $\mathbf{X_i} = \mathbf{Z_i} + \mathbf{X_a}$, we can see that $\mathbf{X_2} - \mathbf{X_1} = \mathbf{Z_2} - \mathbf{Z_1}$, which indicates the change between the two input vectors.

Reconstruction of Original Vectors The original data vectors $\mathbf{X_1}$ and $\mathbf{X_2}$ can be reconstructed as a linear combination of the eigenvector $\mathbf{\Phi} = [\varphi_1\, \varphi_2]^T$ and the coefficient vector $\mathbf{Y} = [y_1\, y_2]^T$ in the transform domain. The zero-mean data matrices $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are reconstructed using $\mathbf{Z}_1 = y_1\mathbf{\Phi}$ and $\mathbf{Z}_2 = y_2\mathbf{\Phi}$, which become:

$$\mathbf{Z}_1 = -(\sqrt{\delta_{11}^2 + \delta_{21}^2}) \begin{bmatrix} \frac{\delta_{11}}{\sqrt{\delta_{11}^2+\delta_{21}^2}} \\ \frac{\delta_{21}}{\sqrt{\delta_{11}^2+\delta_{21}^2}} \end{bmatrix} = \begin{bmatrix} -\delta_{11} \\ -\delta_{21} \end{bmatrix} \tag{7}$$

$$\mathbf{Z}_2 = (\sqrt{\delta_{11}^2 + \delta_{21}^2}) \begin{bmatrix} \frac{\delta_{11}}{\sqrt{\delta_{11}^2+\delta_{21}^2}} \\ \frac{\delta_{21}}{\sqrt{\delta_{11}^2+\delta_{21}^2}} \end{bmatrix} = \begin{bmatrix} \delta_{11} \\ \delta_{21} \end{bmatrix} \quad (8)$$

When the mean vector $\mathbf{X}_a$ is added to the zero-mean vectors $\mathbf{Z_1}$ and $\mathbf{Z_2}$, we obtain the original data vectors $\mathbf{X_1}$ and $\mathbf{X_2}$.

2.2.2 Continuous-Domain Decomposition

Using a continuous multicomponent function $x(t)$ composed of a linear function and a sinusoidal function, we assume that the manufacturing signal is represented with $x(t) = t + sin(wt)$. Detection of linear vectors is the first step in decomposing a general function $g(x,t)$ into its fundamental functions. Linear trends often obscure data; for example, linear trends appear falsely as a low-frequency component using Fourier transform techniques. Typically, pre-processing of the data is necessary to remove the linear trends by means of linear regression techniques.

In the continuous domain, the eigenvector equation for the discrete case is converted to the following integral equation:

$$\int_0^T R(t_1,t_2)\Phi(t_2)dt_2 = \lambda\Phi(t_1) \quad (9)$$

where $R(t_1,t_2) = E[x(t_1)x(t_2)]$ is the covariance function for the real signal $x(t)$, which becomes:

$$\begin{aligned} R(t_1,t_2) &= E[x(t_1)x(t_2)] = E[(t_1 + sin(wt_1))(t_2 + sin(wt_2))] \quad (10) \\ &= E[t_1t_2 + t_1sin(wt_2) + t_2sin(wt_1) + sin(wt_1)sin(wt_2)] \\ &= t_1t_2 + t_1sin(wt_2) + t_2sin(wt_1) + sin(wt_1)sin(wt_2) \end{aligned}$$

The integral equation can be solved by differentiating it with respect to t_1 twice and applying the appropriate boundary conditions, using the Leibnitz Rule. Differentiating the integral equation twice with respect to t_1, we obtain:

$$\int_0^T (-w^2t_2sin(wt_1) - w^2sin(wt_1)sin(wt_2))\phi(t_2)dt_2 = \lambda\phi^{''}(t_1) \quad (11)$$

Adding and subtracting $w^2\int_0^T(t_1t_2 + t_1sin(wt_2))\phi(t_2)dt_2$, the equation becomes:

$$\begin{aligned} &-w^2\int_0^T (t_1t_2 + t_1sin(wt_2) + t_2sin(wt_1) + sin(wt_1)sin(wt_2))\phi(t_2)dt_2 \quad (12) \\ &+w^2\int_0^T (t_1t_2 + t_1sin(wt_2))\phi(t_2)dt_2 = \lambda\phi^{''}(t_1) \\ \Leftrightarrow\quad &\lambda\phi^{''}(t_1) + w^2\lambda\phi(t_1) = w^2\int_0^T (t_1t_2 + t_1sin(wt_2))\phi(t_2)dt_2 \end{aligned}$$

Differentiating both sides of this relationship twice with respect to t_1 once again, we obtain the following differential equation:

$$\frac{d}{d^2t_1}(\lambda\phi^{''}(t_1) + w^2\lambda\phi(t_1)) = \lambda\phi^{''''}(t_1) + w^2\lambda\phi^{''}(t_1) = 0 \quad (13)$$

The fourth order differential equation is solved by first taking the Laplace transform of both sides, and then solving the characteristic equation for its roots as follows:

$$s^4 + w^2 s^2 = 0 \Leftrightarrow s_1, s_2 = 0; s_3, s_4 = +- jw \tag{14}$$

The characteristic equation has two repeated roots, and two complex conjugate roots, resulting in the following general solution for $\phi_i(t)$, which become:

$$\phi_1(t) = A + Bt = Bt \quad and \quad \phi_2(t) = Ccos(wt) + Dsin(wt) = Dsin(wt), \tag{15}$$

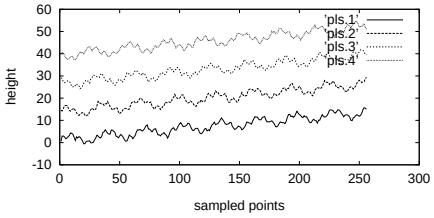
once the boundary conditions are applied at $t = 0$. The resulting eigenfunctions correspond to a linear function and a sinusoidal function. The remaining constants can be found using the boundary condition at $t = T$ and then normalized to satisfy the orthonormality condition.

2.3 Application to Numerical Signals

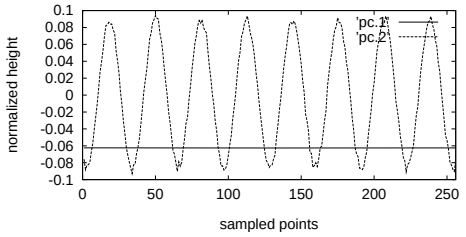
In order to understand the real-world situations, we need to study the decomposition properties of multiple N-dimensional (discretized) signals. In this section, numerically-generated multicomponent signals, composed of two pure sinusoids of different frequency, a linear trend, and random noise, are investigated. These multicomponent signals are representative of signals measured from a manufacturing machine.

For this study, $M = 40$ snapshots (input vectors) are assumed to be collected from a manufacturing process, each with $N = 256$ sampled points, shown in Figure 2(a). The KL transform of the multicomponent signal results in three individual fundamental patterns, shown in Figures 2(b) and 2(c), with the noise component filtered. These fundamental patterns correspond to a low-frequency sinusoid, a high-frequency sinusoid, and a linear function. These three eigenvectors are the coherent components of the original signals, hence demonstrating the decomposability of such complex signals using the KL transform.

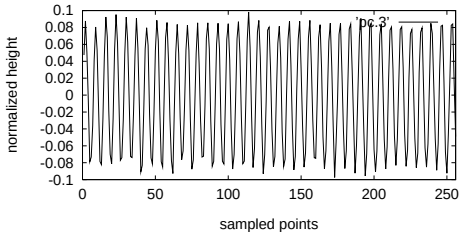
The coefficient vectors corresponding to the three fundamental patterns show the changes in each fundamental pattern over the collected snapshots, as shown in Figures 2(d) and 2(e). Figures 2(b) and 2(c) show the fundamental patterns corresponding to the linear trend, low-frequency sinusoid, and the high-frequency sinusoid. Any changes in the mean or mean-square values of the fundamental patterns (identified by the eigenvectors) can be detected and monitored by means of the corresponding coefficient vectors. For example, Figure 2(d) indicates the change in the slope of the linear trend. The coefficient vectors in Figure 2(e) indicate the relative weight (significance) of each sinusoidal component; any changes in the amplitude of the individual frequency components would be indicated by these coefficient vectors [37]. Examples of further nonstationary changes such as offsets and changes in the variance of the data are provided in [35].



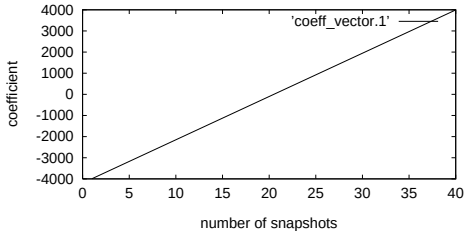
(a) Input Profiles (4 of 40 shown).



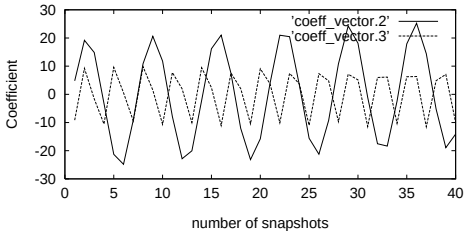
(b) Eigenvectors 1 and 2.



(c) Eigenvector 3.



(d) Coefficient Vector 1.



(e) Coefficient Vector 2 and 3.

Figure 2: Simulation Results: Multicomponent Signals.

2.4 Computational Problems due to Large Dimensionality

The number of sampled data points is an important factor in determining the computational efficiency of the KL transformation. The Karhunen-Loève transform requires the collection of N-dimensional data into a covariance matrix of dimension N by N. The N by N covariance matrix is then resolved to compute the eigenvectors and eigenvalues. If the dimension N is very large, then this requires tremendous computation time, since the transform essentially diagonalizes an N by N matrix. In this section, it is shown that the number of input vectors M is also an important factor in the KL transformation, and can be used to eliminate the problems caused by high-dimensionality in the input data.

When the number of input vectors M is small compared to the data dimension N, the rank of the

covariance matrix is determined by M. This situation is very often encountered in pattern recognition problems, especially if N is very large [14, 34]. In such a case, the covariance matrix is singular and has M or less independent vectors. The covariance matrix results in M eigenvectors only, with the remaining eigenvectors belonging to the null space of this degenerate matrix. For this type of problem, instead of calculating the eigenvalues and eigenvectors from an N by N matrix, the problem can be converted to a much simpler version to compute the eigenvalues and eigenvectors of an M by M matrix. So, instead of using the N by N sample matrix, the eigenvalues and eigenvectors of an M by M matrix are computed using the following eigenvector equation:

$$\frac{1}{M}(\mathbf{U^T U})_{M\times M}\mathbf{\Phi}_{M\times M} = \mathbf{\Phi}_{M\times M}\mathbf{\Lambda}_{M\times M} \tag{16}$$

using the new sample matrix with reduced dimensionality. In this case, an M by M sample matrix $\mathbf{U^T U}$ is diagonalized, which results in M eigenvalues (diagonals of $\mathbf{\Lambda}_{M\times M}$). However, the new eigenvectors are M by M as well, and do not represent the original data correctly, and need to be converted to the actual eigenvectors using:

$$\begin{aligned} & \frac{1}{M}\mathbf{U}_{N\times M}\mathbf{U^T}_{M\times N}\mathbf{U}_{N\times M}\mathbf{\Phi}_{M\times M} = \mathbf{U}_{N\times M}\mathbf{\Phi}_{M\times M}\mathbf{\Lambda}_{M\times M} \\ \Leftrightarrow \quad & \frac{1}{M}(\mathbf{U U^T})_{N\times N}(\mathbf{U}_{N\times M}\mathbf{\Phi}_{M\times M}) = (\mathbf{U}_{N\times M}\mathbf{\Phi}_{M\times M})\mathbf{\Lambda}_{M\times M} \end{aligned} \tag{17}$$

where the product $\mathbf{U}_{N\times M}\mathbf{\Phi}_{M\times M}$ converts the new eigenvectors to their original format, providing a means to transform the M eigenvectors of dimension M in $\mathbf{\Phi}_{M\times M}$ to M eigenvectors of dimension N in $(\mathbf{U}_{N\times M}\mathbf{\Phi}_{M\times M}) = \mathbf{\Phi}_{N\times M}$ (the remaining $N-M$ eigenvectors belong to the null space, since the rank of the covariance matrix is M). This new orthogonal matrix is orthonormalized by dividing the converted eigenvectors by $\sqrt{M\lambda_i}$.

2.5 Accuracy Problems for Noisy Signal Decomposition

The number of input signals is an important factor in determining the accuracy of the decomposition, especially in the presence of highly noisy environments. Consider the goodness measures for estimators, in the context of the covariance matrix estimate. To have an unbiased estimator, the expected value of the covariance estimator has to equal the true value. This is shown as follows:

$$E[\hat{\mathbf{S}}] = E[\frac{1}{M}\sum_{i=1}^{M}\mathbf{X}_i\mathbf{X}_i^T] = \frac{1}{M}E[\sum_{i=1}^{M}\mathbf{X}_i\mathbf{X}_i^T] = \frac{1}{M}\sum_{i=1}^{M}E[\mathbf{X}_i\mathbf{X}_i^T] = \frac{1}{M}\ M\ (\mathbf{S}) = \mathbf{S} \tag{18}$$

Since the inputs are independent, the expected value of the sum becomes the sum of expected values. This shows that the expected value of the covariance estimator is equal to the true value of the covariance, hence resulting in zero bias.

The variance of the covariance estimator is computed as follows:

$$V[\hat{\mathbf{S}}] = V[\frac{1}{M}\sum_{i=1}^{M}\mathbf{X}_i\mathbf{X}_i^T] = (\frac{1}{M})^2\sum_{i=1}^{M}V[\mathbf{X}_i\mathbf{X}_i^T] = \frac{1}{M^2}\ M\ V[\mathbf{X}_i\mathbf{X}_i^T] = \frac{1}{M}\ V[\mathbf{X}_i\mathbf{X}_i^T] \tag{19}$$

Therefore, the variance of the covariance estimator is inversely proportional to the number of realizations M. Hence, as in the case of the power spectral estimator, the more ensemble averages are performed, the smaller is the variance of the covariance estimator, and the more accurate is the estimate of the covariance matrix.

Now, since the covariance matrix estimator has a variance which needs to be reduced to increase the accuracy of the estimation, presumably, so must its eigenvalues and eigenvectors. One could picture this problem as computing the eigenvalues and eigenvectors over small portions of a long sequence of data, and then averaging them over all the portions to get an average estimate for their values.

Let $\hat{\lambda}_i$ be the estimate of the true eigenvalue λ_i, and let $\hat{\phi}_i$ be the estimate of the true eigenvector ϕ_i. Fukunaga [14] presents approximate formulas for the variance the eigenvalue estimator and the mean-square error for the eigenvector estimator. The variance of the eigenvalue estimator becomes:

$$V[\hat{\lambda}_i] \approx E[(\phi_i^T \, \hat{\mathbf{S}} \, \phi_i)^2] - \lambda_i^2 \tag{20}$$

The variation between the eigenvector estimator and the true eigenvector becomes:

$$E[||\hat{\phi}_i - \phi||^2] \approx \sum_{j=1}^{n} \frac{E[(\phi_i^T \, \hat{\mathbf{S}} \, \phi_i)^2]}{(\lambda_i - \lambda_j)^2} \tag{21}$$

where $n < M$ is the number of dominant eigenvalues.

These equations need to be simplified one step further by assuming a distribution type for the expected value operator. Making use of the Central Limit Theorem, Fukunaga assumes that the $\mathbf{X}_i$ are normally distributed, and hence the coefficients are normally distributed. The equations above are then reduced to the following simpler versions:

$$Var[\hat{\lambda}_i] \approx \frac{2}{M}\lambda_i^2 \tag{22}$$

$$E[||\hat{\phi}_i - \phi||^2] \approx \frac{1}{M} \sum_{j=1}^{n} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} \tag{23}$$

3 Example in Manufacturing

In this paper, the focus is on forming an understanding of the mathematical background and demonstrating the mechanics of a fault detection and monitoring technique based on the Karhunen-Loève transform. It is, however, useful to present a simple example of a real-world monitoring situation, using data collected from a manufacturing process. Other examples can be found in [35]. The following is a brief discussion of monitoring the condition of surfaces from a milling process.

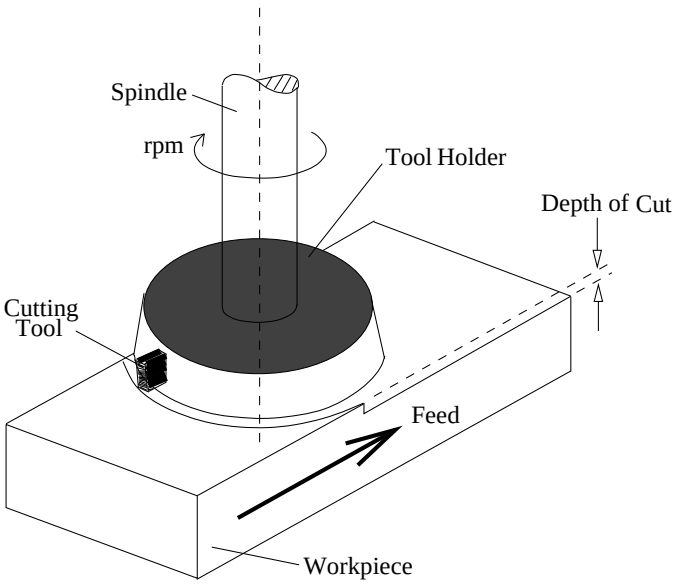


Figure 3: The Face Milling Process.

3.1 Part Surfaces from a Milling Process

To illustrate the use of the method in manufacturing, parts manufactured using a vertical milling machine are used. Vertical mills have an axis of the cutter perpendicular to the workpiece surface. In particular, in face milling, the cutting tooth is attached to the cutter face which is perpendicular to the axis, as shown in Figure 3.

A flat feature, made of Aluminum 2024, has been milled using a vertical milling machine. The cutting tool material is high speed steel. No cutting fluid has been used. $M = 60$ surface profile measurements are collected on the part surface, with $N = 1000$ points in each input vector. These input profiles are sampled at a rate of 799 $points/mm$.

The milling process produces a dominant periodic pattern on part surfaces, due to the feed marks produced during cutting. In addition, nonstationary linear trends appear on the surfaces due to misalignments of the work table, as well as other nonstationarities due to time-varying wear of cutting tool, or sudden chip breakage, etc. Such nonstationarities make it difficult to detect the exact nature of the main periodic component. For example, linear trends often appear as a dominant low-frequency periodic component using Fourier-based methods, which makes the automatic detection of the relevant component difficult. In addition, the nature of the nonstationarities is impossible to determine with Fourier-based methods. As a result, to get more reliable surface information, we analyze such surfaces using our detection technique, based on the KL transform.

3.2 Analysis of Milled Surfaces Using KL

The KL transform of the collected milling profile measurements results in three fundamental eigenvectors, as shown in Figure 4. The corresponding coefficient vectors are shown in Figure 5. At first glance, we observe that the main three components of the milling profiles are a linear trend, a low-frequency sinusoidal pattern, and a high-frequency sinusoidal pattern.

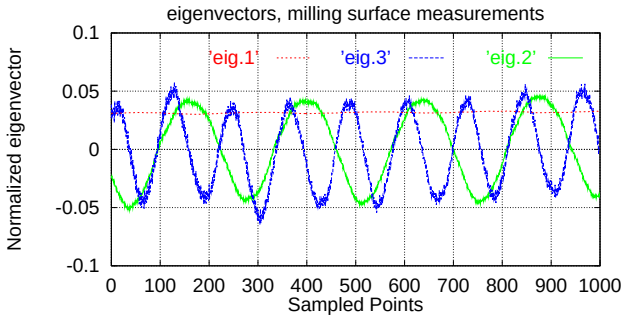
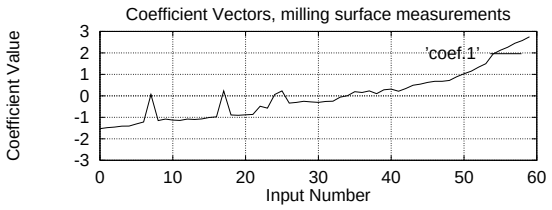
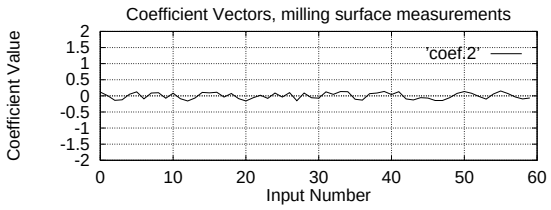


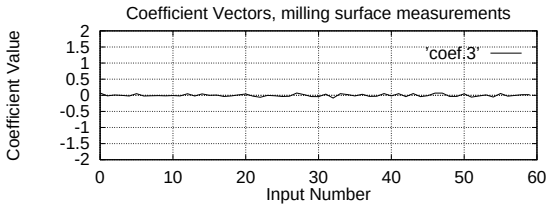
Figure 4: Milling Decomposition: Three Fundamental Modes.



(a) Coefficient Vector 1.



(b) Coefficient Vector 2.



(c) Coefficient Vector 3.

Figure 5: KL Coefficient Vectors for Milled Profiles.

The first eigenvector, shown in Figure 4, corresponds to the linear pattern which exists on the milled surface, indicated by the nature of the eigenvector. Such a trend can be caused by misalignments of the worktable, for example. The corresponding coefficient vector, shown in Figure 5(a), indicates the change in slope and severity of this fundamental pattern over time. The nature of this nonstationary pattern can be

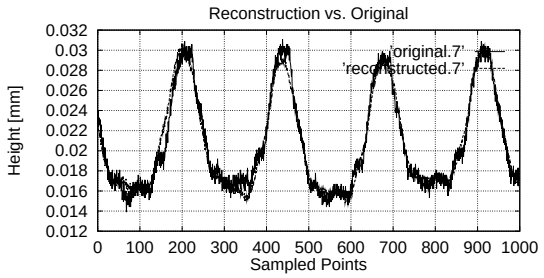


Figure 6: KL Reconstruction of Milled Profiles Using 5 Eigenvectors.

characterized by monitoring the eigenvector and coefficient vector pair. The second mode, shown in Figure 4, corresponds to the main periodic pattern generated on milled surfaces due to the feed marks during milling. A frequency component at $f_1 = 3.19\ cycles/mm$ is clearly identified with the periodic eigenvector, without the adverse effects of nonstationary trends. Any changes in this component, or the relative severity of its magnitude, can be monitored and detected by means of the corresponding coefficient vector, shown in Figure 5(b). The third mode, shown in Figure 4, corresponds to the second frequency component on the surfaces. This additional frequency component at $f_2 = 2f_1 = 6.39\ cycles/mm$, corresponds to the harmonic of the main frequency component due to the feed marks, and is generated due to vibrations of the cutting tool during the spindle rotation.

Figure 6 shows a reconstruction of a profile estimate using the fundamental eigenvectors, compared to the original input profile. Notice that the stochastic component is filtered out of the fundamental eigenvectors, which is one of the advantages of this transform [35]. As a result, the reconstructed estimates have a much less noisy shape. The general shape of the profiles is captured accurately using the first few eigenvectors [35]. This comparison will provide engineers with a means of deciding on the accuracy of the KL analysis results.

4 Conclusions

This paper presents the fundamentals of an alternative manufacturing condition monitoring technique, based on the Karhunen-Loève (KL) transform. First, we establish the mechanics of the technique using simple input vectors. Using these two-dimensional inputs, we also show that the eigenvectors contain the fundamental patterns in the data, and that the coefficient vectors indicate changes in the fundamental patterns over time. Next, we demonstrate the extension to functions with higher dimensionality using a multicomponent input in the continuous domain. Finally, to address the case of general multicomponent inputs in real-world situations, we illustrate the decomposability of high-dimensional discretized inputs using numerically-generated data. The results are then discussed in the context of condition monitoring of real signals from a milling process, monitored by means of the KL-based analysis of surface profiles.

The proposed technique presents a promising alternative to the current condition monitoring techniques. Empirically-computed eigenvectors allow the detection of various types of characteristics, whether linear, sinusoidal, stochastic, or nonstationary. This is an improvement over methods which project the data onto pre-determined functions, such as sines and cosines, as in the case of Fourier-based methods. In addition, the nonstationary nature of the signals is detected and located very effectively by means of the coefficient vectors. Furthermore, a technique based on the KL transform has the potential to offer a clear and unambiguous picture of the signal's content, making its implementation in manufacturing practice more acceptable.

Acknowledgment This material is based on work supported, by the National Science Foundation, Grant No. DDM-9111372; an NSF Young Investigator Award; a research grant from TARP; research grants from Ford Motor Company, Texas Instruments, and Desktop Manufacturing Inc., and the June and Gene Gillis Endowed Faculty Fellowship in Manufacturing.

References

[1] N. Ahmed and K.R. Rao. *Orthogonal Transforms for Digital Signal Processing.* Springer-Verlag, NewYork, 1975.

[2] V.R. Algazi, K.L. Brown, and M.J. Ready. Transform representation of the spectra of acoustic speech segments with applications, Part I: General approach and application to speech recognition. *IEEE Transactions on Speech and Audio Processing*, 1(2):180–195, April 1993.

[3] Gregorio Andria, Mario Savino, and Amerigo Trotta. Application of Wigner-Ville distribution to measurements on transient signals. *IEEE Transactions of Instrumentation and Measurement*, 43(2):187–193, April 1994.

[4] K.S. Ball, L. Sirovich, and L.R. Keefe. Dynamical eigenfunction decomposition of turbulent channel flow. *International Journal for Numerical Methods in Fluids*, 12:585–604, 1991.

[5] R.W. Barker, G. Klutke, and M.J. Hinich. Monitoring rotating tool wear using higher-order spectral features. *Journal of Engineering for Industry*, 115:23–29, February 1993.

[6] J.S. Bendat and A.G. Piersol. *Random Data: Analysis and Measurement Procedures.* John Wiley & Sons, New York, NY, 1986.

[7] J.E. Berry. How to track rolling element bearing health with vibration signature analysis. *Sound and Vibration*, pages 24–35, November 1991.

[8] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time Series Analysis: Forecasting and Control.* Prentice Hall, New Jersey, 1994.

[9] S. Braun. *Mechanical Signature Analysis: Theory and Applications.* Academic Press, London, 1986.

[10] L. Cohen. Time-frequency distributions–a review. In *Proceedings of the IEEE*, volume 77/7, pages 941–981, 1989.

[11] G. Dalpiaz and A. Rivola. Condition monitoring and diagnostics in automatic machines: Comparison of vibration analysis techniques. *Mechanical Systems and Signal Processing*, 11(1):53–73, 1997.

[12] S.D. Eppinger, C.D. Huber, and V.H. Pham. A methodology for manufacturing process signature analysis. *Journal of Manufacturing Systems*, 14(1):20–34, 1995.

[13] J.W.A. Fackrell, P.R. White, J.K. Hammond, and R.J. Pinnington. The interpretation of the bispectra of vibration signals. part I: Theory. *Mechanical Systems and Signal Processing*, 9(3):257–266, 1995.

[14] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, New York, NY, 1990.

[15] Z. Geng and L. Qu. Vibrational diagnosis of machine parts using the wavelet packet technique. *British Journal of Non-Destructive Testing*, 36(1):11–15, January 1994.

[16] M.D. Graham, S.L. Lane, and D. Luss. Proper orthogonal decomposition analysis of spatiotemporal temperature patterns. *Journal of Physical Chemistry*, 97(4):889–894, 1993.

[17] Conor Heneghan, Shyam M. Khanna, and Ake Flock. Investigating the nonlinear dynamics of cellular motion in the inner ear using the short-time fourier and continuous wavelet transforms. *IEEE Transactions on Signal Processing*, 42(12):3335–3351, December 1994.

[18] K. Jemielniak and O. Otman. Catastrophic tool failure detection based on acoustic emission signal analysis. *Annals of the CIRP*, 47(1):31–34, 1998.

[19] R.M. Jones. A guide to the interpretation of machinery vibration measurements, Part I. *Sound and Vibration*, pages 24–35, May 1994.

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

[21] W. Kozek. Matched generalized gabor expansion of nonstationary processes. In *The Twenty Seventh Asilomar Conference on Signals, Systems, & Computers*, volume 1, pages 499–503, 1993.

[22] M.D. Ladd. *Detection of Machinery Faults in Noise Using Wavelet Transform Techniques*. PhD thesis, The University of Texas, Austin, Tx, 1995.

[23] S-H. Lee, H. Zahouani, R. Caterini, and T.G. Mathia. Morphological characterisation of engineered surfaces by wavelet transform. *International Journal of Machine Tools and Manufacture*, 38(5-6):581–589, 1998.

[24] C.J. Li, J. Ma, and B. Hwang. Bearing condition monitoring by pattern recognition based on bicoherence analysis of vibrations. *Proceedings of the Institution of Mechanical Engineers*, 210:277–285, 1996.

[25] E.B. Martin, A.J. Morris, and J. Zhang. Process performance monitoring using multivariate statistical process control. In *IEE Proceedings on Control Theory Applications*, volume 143:2, 1996.

[26] J.S. Mitchell. *Introduction to Machinery Analysis and Monitoring*. PennWell Publishing, Oklahoma, 1993.

[27] P. W. Prickett and C. Johns. An overview of approaches to end milling tool monitoring. *International Journal of Machine Tools and Manufacture*, 39:105–122, 1999.

[28] R.A. Rohrbaugh. Application of time-frequency analysis to machinery condition assessment. In *The twenty-seventh Asilomar Conference on Signals, Systems, & Computers*, volume 2, pages 1455–1458, 1993.

[29] Brian M. Sadler, Georgios B. Giannakis, and Keh-Shin Lii. Estimation and detection in nongaussian noise using higher order statistics. *IEEE Transactions on Signal Processing*, 42(10):2729–2751, December 1994.

[30] L. Sirovich and L.R. Keefe. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America*, 4(3):519–524, March 1987.

[31] J. Sottile and L. E. Holloway. An overview of fault monitoring and diagnosis in mining equipment. *IEEE Transactions on Industry Applications*, 30(5):1326–1332, September/October 1994.

[32] S. Spiewak. A predictive monitoring and diagnosis system for manufacturing. *Annals of the CIRP: manufacturing technology*, 40(1):401–404, 1991.

[33] W. J. Staszewski, K. Worden, and G. R. Tomlinson. Time-frequency analysis in gearbox fault detection using the wigner-ville distribution and pattern recognition. *Mechanical Systems and Signal Processing*, 11(5):673–692, 1997.

[34] C.W. Therrien. *Discrete Random Signals and Statistical Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1992.

[35] I.Y. Tumer. *Foundations of Condition Monitoring for Manufacturing and Design*. PhD thesis, The University of Texas, Austin, Tx, 1998.

[36] I.Y. Tumer, D.C. Thompson, K.L. Wood, and R.H. Crawford. Characterization of surface fault patterns with application to a layered manufacturing process. *Journal of Manufacturing Systems*, 17(1):23–36, January 1998.

[37] I.Y. Tumer, K.L. Wood, and I.J. Busch-Vishniac. Improving manufacturing precision using the Karhunen-Loève transform. In *1997 ASME Design for Manufacturing Conference*, volume DETC97-DFM4347 (cdrom), September 1997.

[38] I.Y. Tumer, K.L. Wood, and I.J. Busch-Vishniac. Monitoring fault condition during manufacturing using the Karhunen-Loève transform. In *1997 ASME Mechanical Vibration and Noise Conference*, volume DETC97-VIB4234 (cdrom), September 1997.

[39] D.J. Whitehouse. *Handbook of Surface Metrology*. Institute of Physics Publishing, Bristol, UK, 1994.

[40] D.J. Whitehouse and K.G. Zheng. The use of dual space-frequency functions in machine tool monitoring. *Measurement Science and Technology*, 3:796–808, 1992.

[41] J.-M. Wu, J.-Y. Lee, and C.-Y. Liou. Diagnoses for machine vibrations based on self organization neural network. In *IECON'91*, pages 1506–1510, 1991.

[42] Y. Zeng and E. Forssberg. Application of vibration signal measurement for monitoring grinding parameters. *Mechanical Systems and Signal Processing*, 8(6):703–713, 1994.

[43] J. Zhang, E. Martin, and J.A. Morris. Fault detection and classification through multivariate statistical techniques. In *Proceedings of the American Control Conference*, pages 751–755, Seattle, Washington, June 1995.