Diagnostic Information Processing for Sensor-Rich Distributed Systems

Elmer S. Hung and Feng Zhao
Xerox Palo Alto Research Center
3333 Coyote Hill Rd.
Palo Alto, CA 94304
{eshung,zhao}@parc.xerox.com

Abstract

This paper describes a diagnostic system for processing high-bandwidth vibration data from distributed sensors for monitoring and diagnosis of electromechanical machines. The system employs time-frequency and principal component analysis techniques to extract and compress features and a Bayesian decision analysis to combine and classify data from multiple sources. Experimental multi-sensor diagnosis results are reported for classifying motor and solenoid vibration signatures from the paper drive plate of the Xerox DC265 digital copier.

Keywords: diagnostics, sensor fusion, time-frequency analysis, wavelets, STFT, Bayesian decision analysis

1 Introduction

Recent advances in batch-fabricated micromachined sensors and electronics have the potential to enable a new generation of condition-based monitoring and diagnostic systems for complex machinery. However, taking advantage of increased sensor and processing capabilities demands corresponding advances in computational techniques for analyzing massive amounts of data from distributed sensor systems.

The work in this paper is part of a larger effort at Xerox to identify and develop scalable processing architectures for interpreting data from many sensors that may be scattered or embedded inside a system. Specifically, in this paper, we present work on combining information from vibration sensors for in-situ diagnosis of component health on a paper drive plate from a Xerox Document Centre 265 (DC265) digital copier. The focus is on sensor data analysis, including the development of flexible time-frequency diagnostic filtering techniques for sensor-rich environments.

Although other researchers (e.g. [1]) have investigated time-frequency techniques for diagnostic vibration analysis, the emphasis in this paper is on the use of time-frequency filters to combine information from large distributed networks of sensors for machine diagnostics. We present a framework for analyzing time-frequency data from many sensors, where it is assumed that training data from lifetime tests is available, but it is too costly to explicitly model the dynamics of the physical environment around every sensor.

The diagnostic data processing system we investigate involves four major components: (1) signal feature extraction, (2) data clustering and compression of high-dimensional feature space information, (3) data aggregation of signal features from multiple sensors, and (4) signal classification and decision analysis. (See the block diagram in Fig. 1.) Data from lifetime tests of the system are used to train the diagnostics algorithm about normal and abnormal operating characteristics. The training is performed offline to generate the run-time diagnostic algorithm.

In the following sections, we first describe the experimental setup for the diagnostics testbed (Section 2). Section 3 describes time-frequency feature extraction techniques, the short-time Fourier transform (STFT) and wavelet analysis. Section 4 describes the use of principal component analysis (PCA) and other techniques to compress the resulting high-dimensional feature space onto a lower dimensional subspace based on the training data from lifetime tests.

The feature space data from different sensors or data analysis methods is aggregated using a statistical Bayesian analysis of probability density functions extracted from the training data (Section 5). A Bayesian discriminant function (Section 6) is then used to produce a simple "health index" that

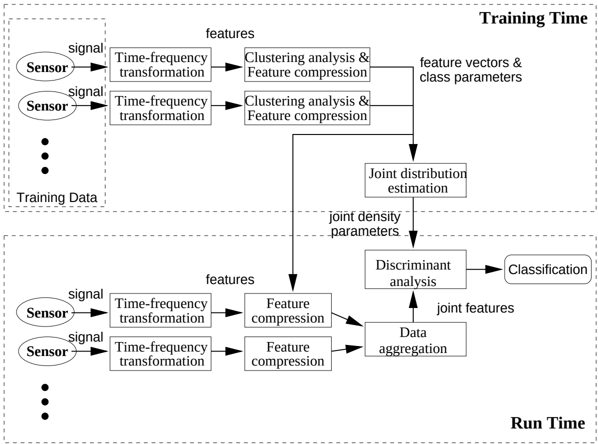


Figure 1: Block diagram of the diagnostic system.

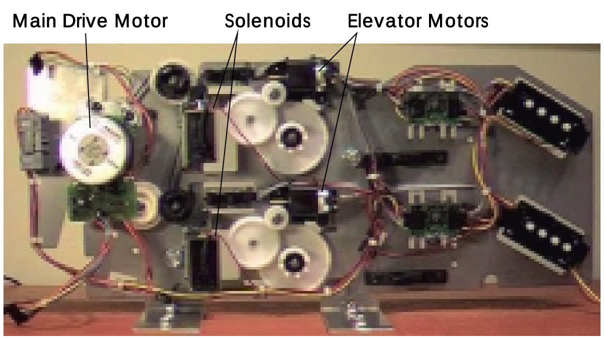


Figure 2: Diagnostic testbed: paper drive plate subsystem of Xerox DC265 copier.

measures how far the system is away from normal operation, or, more generally, an index that measures how close the current behavior is to identified modes of operation or fault conditions. These indices can then be used to generate decision classifications. Finally, we report on the results of applying this algorithm to experimental data (Section 7).

2 Experimental Setup

The drive plate subsystem of a Xerox DC265 copier (see Fig. 2) was chosen as a diagnostics testbed because it is a relatively complex electromechanical system involving different types of actuators with multiple modes of operation. The drive plate is responsible for acquiring paper from the paper tray and directing it to the paper path for xerographic processing. It contains a number of actuators including a main motor, two solenoids, and two elevator motors.

The testbed is instrumented with vibration sensors (PCB Piezotronics, model J352C65). Vibration sensors are used in the initial tests because they provide generic information that is useful for diagnosing different types of actuators, and because they require high data throughputs that may be important in future distributed sensor applications. The vibration sensors on the drive plate are sampled by an analog-to-digital converter (ADC) at 50 kHz with 12-bit resolution. The data is oversampled to provide better accuracy in the optimal bandwidth of the sensor (10 Hz-8 kHz).

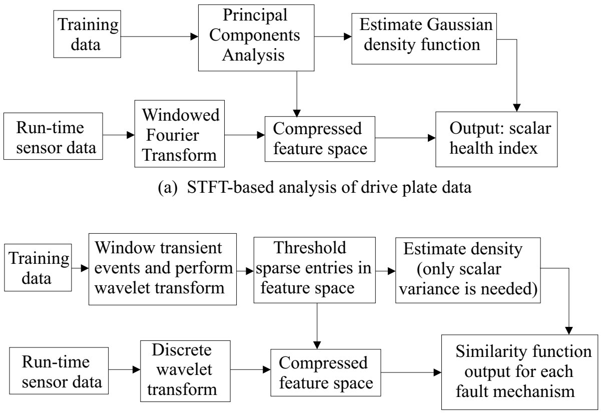
In order to test the diagnostic processing techniques, actuator behavior is purposely compromised in some experiments. In one experiment a washer is attached to the main motor to simulate unbalanced behavior. In another experiment a rubber plug is used to limit the plunger travel distance in one of the solenoids. The objective of these experiments is to see if it is possible to distinguish normal from compromised actuator behavior and to identify actuator operating modes by analyzing the vibration signatures from multiple sensors.

3 Time-Frequency Analysis

Two time-frequency-based techniques are used to analyze the vibration data: windowed short-time Fourier transforms (STFT) and wavelet analysis. For the diagnostic algorithms we have implemented, two training steps are performed offline based on data from lifetime tests in order to establish the normal and/or abnormal operating characteristics of the device. Sections 3 and 4 describe the first training step which involves the use of time-frequency analysis to generate a feature space that properly captures diagnostic information. The second training step (Section 5) is a Bayesian analysis of the data, which involves approximating a Gaussian density function in the product space of the feature spaces for all the sensors involved.

For the first training step, the STFT or wavelet analysis may be used directly to create the feature space, or, more likely, the training data may be used to compress the data onto a lower-dimensional feature space. For example, we use principal component analysis (PCA) to reduce feature space dimensionality.

For reference in the following discussion we present block diagrams of the two specific implementations we used in analyzing DC265 paper drive plate data (Fig. 3). One should note, however, that modules in the two approaches could be interchanged if desired. For example, one could use PCA on the wavelet-based feature space as illus-



(a) STFT-based analysis of drive plate data

(b) Wavelet-based analysis of drive plate data

Figure 3: (a) Block diagram showing STFT-based diagnosis algorithm implemented for DC265 drive plate vibration data. (b) Block diagram showing wavelet-based diagnosis algorithm implemented.

trated with the STFT analysis.

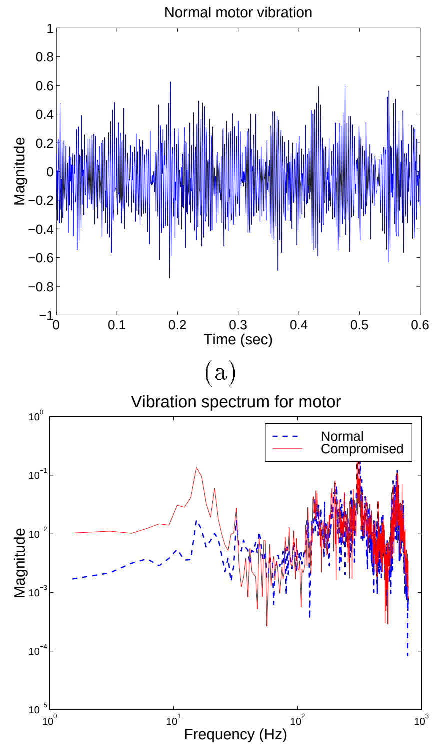
3.1 STFT

STFTs have been used previously in a wide variety of applications including speech, radar, and image processing [2]. For our diagnostic analysis, the strategy can be summed up as follows: During run time, the STFT method involves looking at a sliding window of vibration data in time, computing the Fourier transform of the data in the window, and using the transformed signal to classify the behavior based on its similarity to the training data. The question is how to make the most effective comparison between the incoming data and the training data given data from many sensors.

Note that spectral information from a discrete Fourier transform can be represented in terms of a high dimensional feature space: Suppose that each spectrum has N data points $\{(\omega_1, \alpha_1), (\omega_2, \alpha_2), \ldots, (\omega_N, \alpha_N)\}$, where the ω's are frequency samples and α's are magnitudes of the spectra. Each spectrum can be represented as a single point or vector $(\alpha_1, \alpha_2, \ldots, \alpha_N)$ in an N-dimensional feature space, S.

The first training step involves the collection of spectral data from numerous windows in time. Using principal component analysis, the sample spectra from the training data are used to collapse the high dimensional feature space S onto a lower dimensional subspace that captures the most important diagnostics information.

For example, typical spectra resulting from the main motor vibration signature for a window of length 655 ms is given in Fig. 4 along with the original time-series vibration data. The window length used should depend on the frequency bandwidth of interest. Fig. 4 (b) illustrates a detectable difference in the spectra for normal and compromised motor behavior. Note that for this graph and other vibration analysis in this paper, incoming signals are first decimated and low-pass filtered to reduce high frequency noise and aliasing effects from the sensors.



(a)

(b)

Figure 4: (a) Motor vibration time series (b) Sample frequency spectra for normal vs. compromised motor behavior from a time window of 655 ms.

3.2 Wavelet analysis

A wavelet transformation of a signal produces a time and scale (which correlates to frequency) dependent expansion of the signal [3]. It is particularly useful for the analysis of non-stationary or transitory signals that do not have persistent statistical moments. Specifically, the wavelet analysis employs a family of wavelets, the so-called orthonormal basis functions $\{\psi_{u_n,s_j}\}_{(n,j)\in Z^2}$, where u_n and s_j are position and scale parameters and

$$\psi_{u_n,s_j}(x) = |s_j|^{\frac{1}{2}}\psi(\frac{x-u_n}{s_j}).$$

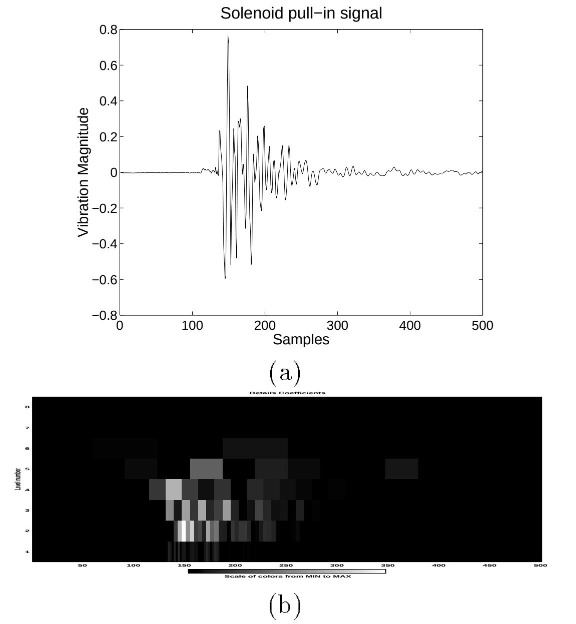


Figure 5: Wavelet analysis of transitory signals: (a) Solenoid pull-in signal; (b) Wavelet coefficients plotted as a two-dimensional intensity map: horizontal axis – time; vertical axis – scale. The coefficients are computed by a discrete wavelet transformation (DWT) with the Daubechies wavelet (db4-8).

Wavelet coefficients are obtained by convolving a wavelet with the signal $f(t)$:

$$a_{u,s} = \int_{-\infty}^{+\infty} f(t)\psi^*_{u,s}(t)dt.$$

Wavelet analysis has been applied by other researchers for fault detection; however most existing approaches aim at novelty or event detection against a steady-state background [4, 5], or are tuned to exploit system-specific features [6]. The wavelet-based discriminant analysis presented in this paper exploits not only event detection but also uses specific wavelet patterns to classify signals into different operating conditions.

For example, in Fig. 5 we take a time window of the vibration signal collected from the pull-in actuation of a solenoid on the testbed. The result is illustrated as an intensity map in the two-dimensional time-scale plane, the horizontal axis is the time, and vertical axis is the scale of the signal. Each tile records the amplitude of the wavelet expansion coefficient for the signal at scale level 2^j and position $2^j n$ in time.

In Fig. 5(b), the onset of the pull-in signal can be identified in time by a set of high-intensity tiles at levels 1-4. As features of the signal vary, the wavelet coefficients vary accordingly. Thus, the intensity variation in the distribution map of the wavelet coefficients could be used to pin-point when a change of interest occurs and what the change is in monitoring and diagnosis.

During the training phase, multiple windows $\{s_1, s_2, ..., s_n\}$ of transitory events (in our case solenoid firings) are extracted to form a training population for a given operating mode or condition. A feature vector $\vec{a_i} \in S$ for window s_i is formed by concatenating the coefficients of the wavelet transformation of s_i at all levels of interest. In practice, the number of levels to consider is a function of the signal energy distribution across the levels, also to be determined from the training samples.

4 Feature space compression

The feature space S produced by the STFT or wavelet transform is often high dimensional. This section describes techniques to compress this feature space into a more manageable lower-dimensional space. This is important in order to alleviate problems from overfitting. It also helps reduce communications and computational bandwidth requirements which is potentially critical for large sensor networks where data from many sensors must be aggregated.

4.1 PCA

For the STFT algorithm, principal component analysis (PCA) was used to project S onto a lower $k-$dimensional subspace, S_k. In the past, PCA has been used in a wide variety of settings including as a feature-space compression technique for a Bayesian classifier [7], and to produce reduced data characterizations of STFT's [8]. It seems particularly well-suited to the sensor fusion problems addressed in this paper because of the strong tendency for dimensionality explosion with increasing numbers of sensors.

The principal component analysis can be interpreted as follows: Let $\{x_1, x_2, \ldots, x_n\}$ be vectors in S, each representing a single windowed spectrum from the training data, and let $m = dim(S)$. For any $k < m$ the idea is to project the feature space onto a k-dimensional subspace, S_k, such that S_k minimizes the sum of the squares of the distances from each of the x_i's to S_k.

Computationally, this is accomplished as follows: Let X be the matrix whose columns are $\{x_1, x_2, \ldots, x_n\}$. For each $k < m$, S_k is the space spanned by the principal component vectors, $\{u_1, u_2, \ldots, u_k\}$, which are given by the first k columns of the matrix U in the singular value decomposition (SVD) of X :

$$X = U \Sigma V^T$$

where U and V are orthogonal, and $\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_m)$ with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$, and $p = min(m, n)$.

For $k << N$ the resulting spectral data representations on S_k contain far less data, but should capture the diagnostically useful information. Each x_i is now written in the new k-dimensional coordinate system based on the principal component basis vectors $\{u_1, u_2, \ldots, u_k\}$. For diagnostics purposes, we also keep one additional coordinate for each x_i containing the residual distance between x_i and S_k. This way the resulting feature space not only contains information that closely reconstructs the original feature space, but also provides information about how far away the original data was to the new representation.

4.2 Sparsity of feature space

Alternatively, the sparsity of the feature space can be exploited to compress the dimensions. Feature vectors computed from wavelet transform are typically sparse, with a large number of small coefficients which may be eliminated without losing diagnostically significant information. In other words, a signal can be adequately approximated using only a subset of feature dimensions that are tuned to record larger components of the signal. This technique was used to obtain a factor of two compression ratio for the wavelet analysis of the solenoid pull-in test case.

5 Bayesian Aggregation

Once a suitable feature space representation of the data is obtained, the question is how to combine information from multiple sensors and analysis algorithms, and how to make decisions regarding feature space output.

The first step involves combining features from multiple signals by considering a composite feature space consisting of the product of all the features spaces. In other words, suppose that the feature vectors from m sensors or data sources are given by $\{x_1, x_2, \ldots, x_m\}$. We then consider the composite feature vector $x = \prod_i^m x_i$.

Bayesian decision theory (e.g., [9]) can then be used to aggregate the data in the composite space. The probability that a hypothesis H_i is true given evidence x is given by:

$$P(H_i|x) = \frac{p(x|H_i)P(H_i)}{p(x)}$$

Omitting the normalizing constant $p(x)$ (independent of H_i), the heart of the analysis is to estimate the conditional probability density function $p(x|H_i)$ for each of the hypotheses, H_i.

Note that if data is not available from all the various fault conditions, it is also possible to use this framework to make decisions based on how far from normal the observed behavior is. In this case, the objective would be to simply estimate $p(x|H)$ from the data where H assumes a normal operating condition.

So how does one determine $p(x|H_i)$ from the training data information? In many interesting cases, it is feasible to assume that the sample data distribution $p(x|H_i)$ is close to Gaussian normal. In this situation, we may approximate the density function by simply estimating the mean and covariance of each sample cluster corresponding to a given hypothesis.

There are a few issues to point out. First of all, it may not be possible to approximate the density as Gaussian. In this situation, a number of techniques are available for approximating the density function, such as using a mixture of Gaussians [11], although, of course, the situation becomes more complex if the density function is not easily represented.

Second, if the dimensionality of the original feature space vectors has not been compressed, the dimensionality of the product space of feature vectors may be very large. There may not be enough data to generate a full representation of the density function. This is the case for the wavelet-based analysis (see the discussion in Section 6).

Note that if there are large numbers of sensors, this may also contribute to feature space dimension explosion. In this situation, it may be possible to hierarchically combine data from groups of sensors at a time. This increases the efficiency of the approach and reduces demand on data, but the drawback is that information about cross-correlations between sensor readings may be lost.

6 Discriminant Analysis

In this section, we assume that the density function $p(x|H_i)$ can be represented as a Gaussian. For simplicity, it is then helpful to employ a discriminant function for classifying inputs [10].

We consider the discriminant function $g_i(x)$, a monotonic function of the *a posteriori* probability, $\log P(H_i|x)$:

$$g_i(x) = \log p(x|H_i) + \log P(H_i).$$

Since it is often difficult to estimate the prior $P(H_i)$, for our experiments we drop the $\log P(H_i)$ term and use the $g_i(x) = \log p(x|H_i)$. This gives the maximum likelihood classifier that selects H_i as the winning hypothesis if $\sup(g_i(x)) > \sup(g_j(x))$ for all $j \neq i$, where x may vary over the entire feature space.

For Gaussian density functions we have:

$$p(x|H_i) = \frac{\exp(-\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i))}{(2\pi)^{d/2}|\Sigma_i|^{1/2}}$$

where μ_i and Σ_i are the mean and covariance of feature vectors of class i and d is the dimension of the feature space. Discarding terms in $g(x)$ that is independent of i, we have

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i) - \frac{1}{2}\log|\Sigma_i|. \quad (1)$$

When training data is only available for normal operation (hypothesis H), then $g(x) = p(x|H)$ represents the key statistic to measure. We define a "health index," $h(x)$ as follows:

$$h(x) = ((x-\mu)^t \Sigma^{-1}(x-\mu))^{1/2}$$

$h(x)$ is sometimes called the *Mahalanobis distance* [10], and measures how far x is away from normal behavior. This is the primary statistic used in the STFT analysis.

Note, however, that the covariance matrix Σ_i in Eq. (1) is an $N \times N$ matrix where N is the dimension of the composite feature space. Thus determining Σ_i requires the estimation of $(N^2+N)/2$ scalars. This can lead to overfitting of data if the feature space dimension is large, and the sample pool is small.

For feature spaces compressed with PCA, the full covariance matrix is estimated directly in order to compute $h(x)$. However, for the wavelet analysis, if the feature space is not compressed, then data overfitting is an issue. One way to circumvent this difficulty is to assume $\Sigma_i = \sigma^2 I$ where σ^2 is the Euclidean-norm variance of the feature vectors from the training data.

In this case, the discriminant function $g_i(x)$ simplifies to the following "similarity" measure:

$$s_i(x) = -\frac{|x-\mu_i|^2}{2\sigma^2} - \log\sigma \quad (2)$$

This function measures how close the input signal is to each training sample cluster. The surfaces of constant distance are hyperspheres as opposed to the hyperellipsoids measured by $h(x)$. In practice this means that each feature space or sensor reading is considered independently and cross-correlations are ignored when $s_i(x)$ is used; however the technique is robust for high dimensional spaces since only one variance parameter is estimated.

Alternatively, the approach in [12] does not assume feature independence and instead uses a probabilistic graph to model the dependencies among different levels at additional computational cost.

7 Results

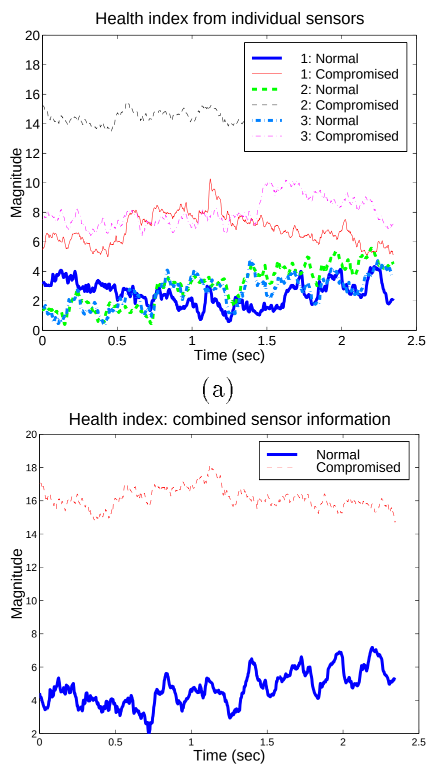
This section describes results for using the diagnostics techniques on data from the DC265 paper drive plate testbed.

Fig. 6 shows the result of applying the STFT-based analysis outlined in Fig. 3(a) onto main motor vibration data from 3 sensors mounted on different parts of the drive plate. Training data is only used for the "normal" motor vibration, and the objective is to see if the resulting health index can distinguish between normal and compromised motor behavior.

A window of length 655 ms is used. Longer windows capture lower frequencies better but are more computationally intensive and take longer to respond. The first 2 principal component features and one residual component feature are kept during the PCA compression stage of the algorithm. Keeping more principal components keeps more information but encourages dimensionality growth. Both parameters may be chosen during training by looking at signal-to-noise ratios as described below.

Fig. 6(a) shows the individual time-series trace of the health index from each of the three sensors. There are six traces corresponding to the results for each sensor on two vibration data sets: one with a normal and one with a compromised motor running. Sensor #2 is the most sensitive to the behavior difference but all three sensors show a separation between the responses of normal and compromised motors (note that smaller values of the health index indicate input closer to normal). Fig. 6(b) shows the composite health index when data from all three sensors are aggregated together. It is clear that a simple thresholding would perform well for selecting normal from compromised behavior.

The composite index performs better than any of



(a)

(b)

Figure 6: (a) Health index for three vibration sensors run on two different data sets: when the normal main motor and when the compromised main motor is running. (b) Composite health index of all three sensors for the same two data set.

the individual sensors. This can be shown by considering signal-to-noise (S/N) ratio statistics. For two hypotheses H_1 and H_2, we define the S/N ratio $r(x)$ of health index $x(t)$ as follows:

$$r(x) = \frac{|E(x|H_1) - E(x|H_2)|}{\sqrt{0.5(Var(x|H_1) + Var(x|H_2))}}$$

where $E(x|H_i)$ is the expected value of x given H_i and $Var(x|H_i)$ is the variance statistic of x given H_i. Table 1 details not only the S/N ratio contribution for each sensor, but also breaks down statistics for each principal component in the sensor feature spaces. Normally one would expect that additional information would tend to improve S/N ratio. Indeed, the composite S/N ratio is higher than that of any single sensor; however, because of the limited sample size of the data, this is not always the case as with the principal component breakdown for sensor #2.

Fig. 7 shows the time-series trace of the health statistic from the same three vibration sensors used in the previous experiment except this time the input consists of pull-in firings from a solenoid. Virtually the same algorithm is used for detecting solenoid behavior as for with the motor except that the window length used is a factor of 4 less (164 ms), since low frequencies are less important for solenoids, and 3 principal components are kept instead of 2.

	component			total
	1	2	3+	
sensor 1	1.79	0.81	4.00	4.69
sensor 2	4.88	0.75	13.60	11.48
sensor 3	2.93	0.74	5.01	5.38
composite				13.56

Table 1: Signal-to-noise ratio results from multi-sensor STFT analysis for distinguishing normal vs. compromised main motor vibration.

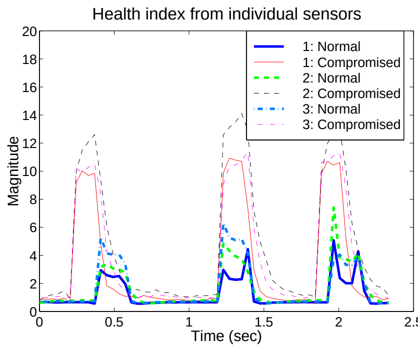
Fig. 7(a) shows 6 time-series of health index traces. Three traces show individual sensor health index results from an experiment with a normal solenoid, three traces show results from a different experiment with an abnormal solenoid. Fig. 7(b) shows composite health index results for the two experiments. We see that the STFT method is flexible enough to detect normal vs. compromised behavior for solenoids as well as motors.

The final two figures show results from the wavelet-based analysis outlined in Fig. 3(b). In this case, we assume training data is available from all the various operating conditions, and the problem is to choose which operating condition or fault hypothesis is correct given new input data.

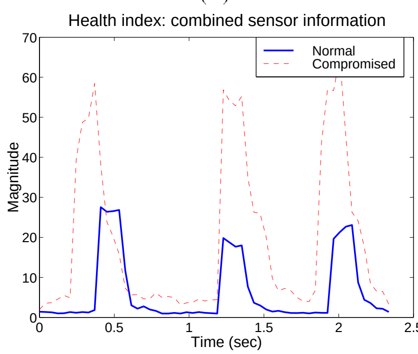
Wavelet coefficient features are extracted from the training data for solenoid and motor vibration signals and compressed using the thresholding technique. We compute the coefficients using the discrete wavelet transform with the Daubechies db5-3 basis functions on windowed signals of length 40 ms.

For each operating condition training data set, the resulting feature space data is used to generate the similarity measure given in Eq. (2). Figs. 8 and 9 plot the similarity measure for signals with respect to each of the known conditions or faults. Using the output of the similarity measure, the system classifies the signal at each time into one of the conditions or faults based on the classification of the dominant response.

In Fig 8, the objective is to classify vibration data from a solenoid pull-in event. The similarity measure exhibits a dominant peak response for the solenoid pull-in condition (solid line), indicat-

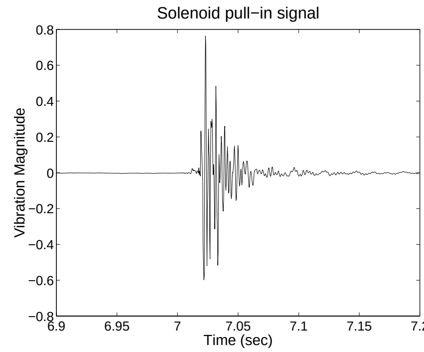


(a)

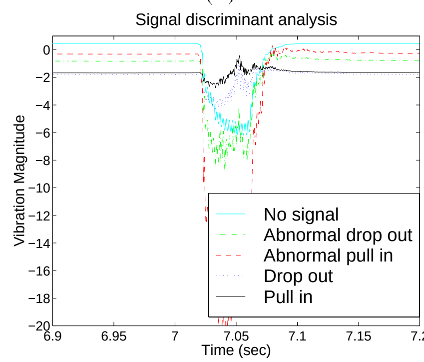


(b)

Figure 7: (a) Health index for three vibration sensors run on two different data sets: pull-in vibration data from a normal and compromised solenoid. (b) Composite health index of all three sensors for the same two data sets.



(a)



(b)

Figure 8: Wavelet-based signal discriminant analysis of solenoid signals: (a) Solenoid pull-in signal; (b) Similarity measures for five conditions: normal pull in, normal drop out, abnormal pull in, abnormal drop out, and no signal.

ing that a pull-in condition has occurred (the time for the event is determined by subtracting the moving window length from the time the peak response in the similarity measure occurs).

In Fig. 9, a solenoid pull-in signal is mixed with a motor signal. In this case, the similarity measure for the pull-in condition also responded with a dominating peak. Note that after the solenoid transient tapers off, the measure for the motor becomes dominant.

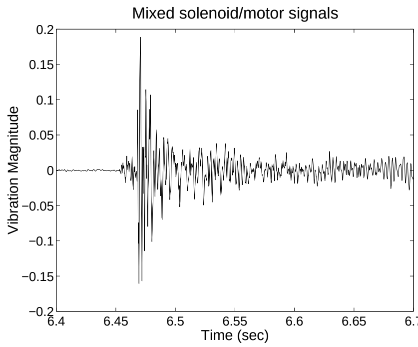
8 Conclusion

In this paper we demonstrate how time-frequency analysis of multiple sensors can be used to diagnose actuator behavior based on vibration data from a complex multi-mode electromechanical system, the Xerox DC265 paper drive plate.

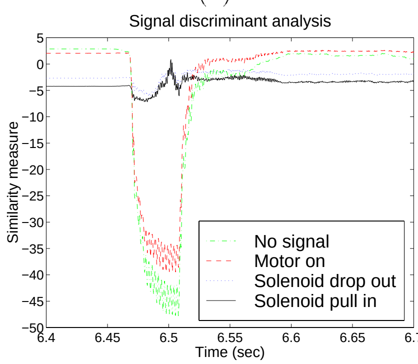
The general objective of this work is to develop scalable processing techniques that are flexible enough to perform on a wide class of distributed sensor and actuator systems without high application-specific overhead. The time-frequency sensor data aggregation techniques in this paper appear to be promising tools, since multi-sensor diagnosis results are demonstrated using radically different solenoid and motor vibration signals with little adjustment in the algorithms. In the case of the wavelet technique, five different fault conditions are reliably classified.

9 Acknowledgment

The authors would like to thank Eric Manders for instrumenting the DC265 drive plate testbed. We would also like to thank Bob Siegel and Charles Coleman for providing the drive plates and helpful discussion about diagnostics techniques. In addition, we thank Eric Manders, Eric Jackson, Koenraad Van Schuylenbergh, Sriram Narasimhan, John Gilbert, and Gautam Biswas for testbed setup assistance, data collection, and useful discussions about algorithms.



(a)



(b)

Figure 9: Wavelet-based signal discriminant analysis of mixed motor and solenoid signals: (a) Solenoid pull-in with motor turned on; (b) Similarity measures for four conditions: solenoid pull in, solenoid drop out, motor on, and no signal.

References

[1] B. Samimy and G. Rizzoni, "Mechanical signature analysis using time-frequency signal" *Proceedings of the IEEE*, 84(9), 1996, 1330-1343.

[2] A.V. Oppenheim and R.W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, 1989.

[3] S. Mallat, *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[4] R.J. Feriez and D.C. Lang, "Gear-tooth fault detection and tracking using the wavelet transform," *Proc. Prognosis of Residual Life of Machinery and Structures*, MFPT, 1998.

[5] Y. Wang, "Jump and sharp cusp detection by wavelets," *Biometrika,* 82(2):385-97, 1995.

[6] K. Bossley, R. Mckendrick, C.J. Harris, and C. Mercer, "Wavelet packet analysis in the condition monitoring of rotating machinery," *Proc. of Symp. on AI in Real-Time Control*, Pergman Press, 1998.

[7] C. Liu and H. Wechsler, "A unified Bayesian framework for face recognition," *Proc. ICIP '98*, v. 1, 1998, 151-155.

[8] D. Beyerback and H. Nawab, "Principal components analysis of the short-time Fourier transform," *Proc. ICASSP '91*, v. 3, 1991, 1725-1728.

[9] J. Manyika and H. Durrant-Whyte, *Data Fusion and Sensor Management*, Ellis Horwood, 1994.

[10] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.

[11] R.A. Redner and H.F. Walker, "Mixture densities, maximum likelihood and the EM Algorithm," *SIAM Review*, 26(2), 1984, 195-239.

[12] M.S. Crouse, R.D. Nowak, and R.G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models." *IEEE Trans. Signal Processing*, 46(4), 1998.