



Intelligent on-line quality control of washing machines using discrete wavelet analysis features and likelihood classification

S. Goumas^a, M. Zervakis^a, A. Pouliezos^{b,*}, G.S. Stavrakakis^a

^aDepartment of Electronic and Computer Engineering, Technical University of Crete, Kounoupidiana, 73100 Chania, Crete, Greece

^bDepartment of Production and Management Engineering, Technical University of Crete, Kounoupidiana, 73100 Chania, Crete, Greece

Received 1 November 1999; received in revised form 1 November 2000; accepted 1 October 2001

Abstract

This paper presents a method for extracting features in the wavelet domain from the vibration velocity signals of washing machines, focusing on the transient (non-stationary) part of the signal. These features are then used for classification of the state (acceptable-faulty) of the machine. The performance of this feature set is compared to features obtained through standard Fourier analysis of the steady-state (stationary) part of the vibration signal. Minimum distance Bayes classifiers are used for classification purposes. Measurements from a variety of defective/non-defective washing machines taken in the laboratory as well as from the production line are used to illustrate the applicability of the proposed method. © 2002 Published by Elsevier Science Ltd.

Keywords: Quality control; Wavelets; Classification; Intelligent systems

1. Introduction

Conventional Fourier analysis provides averaged spectral coefficients which are independent of time (Ambardar, 1995). They represent the frequency composition of a random process which is assumed to be stationary. However, many random processes are essentially non-stationary (Sólnes, 1997). For example, the sound pressure recorded from speech and music is non-stationary (Qian and Chen, 1996); in vibration monitoring, the occurrence of transient impulses makes the recorded signal non-stationary (Newland, 1994a, b; Tamaki et al., 1994; Wang and McFadden, 1994; Wilkinson and Cox, 1996); vibration during the start-up of an engine is non-stationary (Kim et al., 1995), and so on.

The basis functions used in Fourier analysis, sine waves and cosine waves, are precisely located in frequency, but their duration spans the entire time axis. The frequency information of a signal calculated by the classical Fourier transform is an average over the entire

duration of the signal. Thus, if there exists a local transient over some small interval of time in the lifetime of the signal, the transient will contribute to the Fourier transform but its location on the time axis will be lost (Saito, 1994). Although the short-time Fourier transform (Qian and Chen, 1996) overcomes the time location problem to a large extent, it does not provide multiple resolution in time and frequency, which is an important characteristic for analyzing transient signals containing both high- and low-frequency components (Lee and Schwartz, 1995; Qian and Chen, 1996).

Wavelet analysis overcomes the limitations of Fourier methods by employing analysis functions that are local both in time and in frequency (Galli et al., 1996; Vetterli and Kovacevic, 1995). These wavelet functions are generated in the form of translations and dilations of a fixed function, the so-called mother wavelet. The focus of this paper is to present the basic ideas of discrete wavelet analysis and to demonstrate the application of wavelet analysis for feature extraction, in conjunction with statistical digital signal processing techniques (Hayes, 1996; Krauss et al., 1994), to the problem of classification of the state of washing machines based on vibration velocity transient signals.

*Corresponding author. Tel.: +30-821-037313; fax: +30-821-037364.

E-mail address: pouliezo@dssl.tuc.gr (A. Pouliezos).

2. Basic ideas of wavelet analysis

2.1. Wavelet analysis

Wavelet analysis breaks up a signal into shifted and scaled versions of the original (or *mother*) wavelet (Saito, 1994). The analyzing (mother) wavelet determines the shape of the components of the decomposed signal. Wavelets must be oscillatory, must decay quickly to zero, and must have an average value of zero. In addition, for the discrete wavelet transform considered here, the wavelets must be orthogonal to each other.

There are several families of wavelets such as Haar wavelets, Daubechies wavelets, biorthogonal, Coifflets, etc. (Misity et al., 1996; Strang and Nguyen, 1996). The Daubechies family is often represented by DN , where N is the order, or the size of the mother wavelet, and D stands for the “family” of wavelets. This family has been used extensively, since the maximum of the signal energy is contained in a limited number of coefficients in Daubechies wavelets. In this work the D4 wavelet is used, which captures well the characteristics of the vibration velocity signal.

2.2. Scaling functions and wavelet functions

The *dilation equations* may be used to generate orthogonal wavelets. The *scaling function* $\phi(t)$ is a dilated (horizontally expanded) version of $\phi(2t)$. The dilation equation in general has the form:

$$\begin{aligned} \phi(t) = c_0\phi(2t) + c_1\phi(2t - 1) \\ + c_2\phi(2t - 2) + c_3\phi(2t - 3). \end{aligned} \tag{1}$$

For the Daubechies D4 wavelet its coefficients have values:

$$c_0 = (1 + \sqrt{3})/4\sqrt{2}, \tag{2}$$

$$c_1 = (3 + \sqrt{3})/4\sqrt{2}, \tag{3}$$

$$c_2 = (3 - \sqrt{3})/4\sqrt{2}, \tag{4}$$

$$c_3 = -(\sqrt{3} - 1)/4\sqrt{2}, \tag{5}$$

Thus, a particular family of wavelets is specified by a particular set of numbers, called the *wavelet filter coefficients*. The above set of numbers c_0, c_1, c_2, c_3 is called the *D4 wavelet filter coefficients*.

It is not possible in general to solve directly for $\phi(t)$; the obvious approach is to solve for $\phi(t)$ iteratively so that $\phi_j(t)$ approaches $\phi_{j-1}(t)$, where,

$$\begin{aligned} \phi_j(t) = c_0\phi_{j-1}(2t) + c_1\phi_{j-1}(2t - 1) \\ + c_2\phi_{j-1}(2t - 2) + c_3\phi_{j-1}(2t - 3). \end{aligned} \tag{6}$$

Fig. 1 shows the scaling function for the D4 wavelet that is obtained from this iteration process, assuming

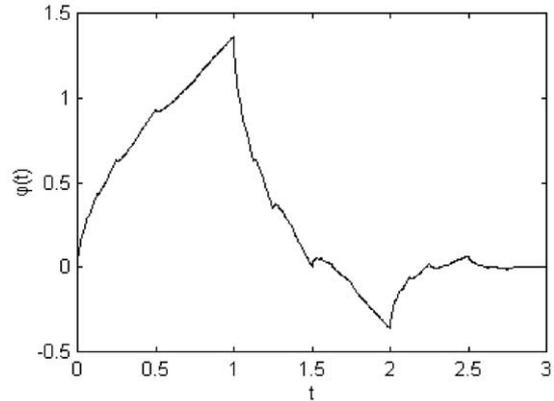


Fig. 1. Scaling function $\phi(t)$ for the D4 wavelet after 8 iterations.

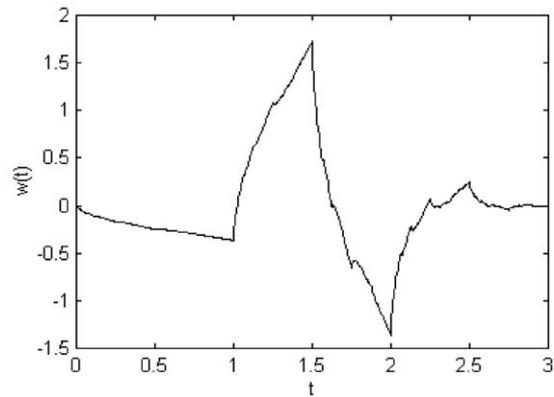


Fig. 2. D4 wavelet function after 8 iterations.

the initial scaling function $\phi_0(t)$ equals 1 for $0 \leq t < 1$ and 0 elsewhere.

The D4 wavelet function $w(t)$ for the four-coefficient scaling function defined in (1) can be computed as

$$\begin{aligned} w(t) = -c_3\phi(2t) + c_2\phi(2t - 1) \\ - c_1\phi(2t - 2) + c_0\phi(2t - 3) \end{aligned} \tag{7}$$

and is shown in Fig. 2.

In general, for an even number M of wavelet filter coefficients $c_k, k = 1, \dots, M - 1$, the scaling function is defined by

$$\phi(t) = \sum_{k=1}^{M-1} c_k\phi(2t - k) \tag{8}$$

and the corresponding wavelet is derived as

$$w(t) = \sum_{k=1}^{M-1} (-1)^k c_k\phi(2t + k - M + 1). \tag{9}$$

It is observed that the scaling function, viewed as a filter’s impulse response, has a low-pass form, whereas the wavelet function has a high-pass form. Thus, the wavelet function is essentially responsible for extracting

the detail (high-frequency components) of the original signal.

2.3. Continuous wavelet transform

The *continuous wavelet transform* (CWT) of a signal $s(t)$ (Strang and Nguyen, 1996) is defined as the integral over time of $s(t)$ multiplied by the scaled and shifted versions of the wavelet function $w(t)$:

$$\text{CWT}(a, b) = \frac{1}{\sqrt{|a|}} \int s(t)w\left(\frac{t-b}{a}\right) dt; \quad a \neq 0. \quad (10)$$

The parameter α represents the scale index that is the reciprocal of the frequency. The parameter b indicates the time shifting (or translation). Suppose that $w(t)$ is centered at time zero and frequency ω_0 . Recall that this signal is highly concentrated both in time and frequency. Then, its dilation and translation $w(a^{-1}(t-b))$ is centered at time b and frequency ω_0/α , respectively. Consequently, the transform $\text{CWT}(a, b)$, as inner product of $s(t)$ and $w(a^{-1}(t-b))$, reflects the signal's behavior in the vicinity of $(b, \omega_0/\alpha)$. Therefore it could also be thought of as a function of time and frequency as $\text{CWT}(a, b)|_{a=(\omega_0/\omega), b=t} = \text{CWT}((\omega_0/\omega), t)$.

The result of the CWT is a function of the *wavelet coefficients* $\text{CWT}(a, b)$, which is a function of scale and position. Multiplying each value $\text{CWT}(a, b)$ by the value $w((t-b)/a)$ yields the portion of the signal $s(t)$ at the corresponding scale and position parameters (a, b) . The exact reconstruction wavelets allow the perfect reconstruction of the original signal $s(t)$. In this case the wavelet function $w(t)$ has to satisfy the *admissibility condition* given by

$$C_W = \frac{1}{2\pi} \int \frac{|W(\omega)|^2}{|\omega|} d\omega < \infty, \quad (12)$$

where $W(\omega)$ is the Fourier transform of the wavelet function $w(t)$. Condition (12) implies $W(\omega) = 0$. In other words, the wavelet function has a *bandpass* behavior. Once $w(t)$ meets the admissibility condition, the original signal $s(t)$ can be reconstructed from

$$s(t) = \frac{1}{C_W} \iint \frac{1}{a^2} \text{CWT}(a, b)w\left(\frac{t-b}{a}\right) da db. \quad (13)$$

Hence, the product $\text{CWT}(a, b)w((t-b)/a)$ is often referred to as the reconstructed signal at scale a and position b .

2.4. Discrete wavelet transform

Calculating the wavelet coefficients as a continuous function of scale and translation is quite complicated in general. It turns out that if scales and positions are chosen based on powers of two in a *dyadic* structure then the analysis becomes much more efficient and just

as accurate. Thus, the *discrete wavelet transform* (DWT) is defined as (Qian and Chen, 1996)

$$\begin{aligned} \text{DWT}(j, k) &= \text{CWT}(a, b)|_{a=2^j, b=k2^j} \\ &= 2^j \int s(t)w(2^j t - k) dt \end{aligned} \quad (14)$$

for $j \in Z, k \in Z$.

The DWT allows a signal $s(t)$ to be decomposed into a series of wavelet coefficients. Using these coefficients, one can exactly reconstruct the original signal as

$$s(t) = \sum_{j \in Z} \sum_{k \in Z} \text{DWT}(j, k)w_{j,k}(t), \quad (15)$$

where $w_{j,k}(t) = w(2^j t - k)$.

The *wavelet coefficients* $\text{DWT}(j, k)$ represent the amplitudes of the contributing wavelets in a similar manner as that the Fourier series coefficients represent the amplitudes of the various sine and cosine terms in the classical Fourier analysis.

2.5. Details and approximations

Unlike conventional techniques, the wavelet analysis produces a family of hierarchically organized decompositions. The selection of a suitable level for the hierarchy depends on the signal and the task to be performed. Often the level is chosen based on a desired low-pass cutoff frequency.

At each level j , is built the j -level approximation, A_j , or the approximation at level j , and a deviation signal called the j -level detail, D_j , or the detail at level j (Misy et al., 1996). The original signal could be considered as the approximation at level 0, denoted by A_0 . The words “approximation” and “detail” are justified by the fact that A_1 is an approximation of A_0 taking into account the “low frequencies” of A_0 , whereas the detail D_1 corresponds to the “high frequency” correction. The organizing parameter, the scale a , is related to level j by $a = 2^j$. If resolution is defined as $1/a$, then the resolution increases as the scale decreases. The greater the resolution, the smaller and finer are the details that can be accessed.

The decomposition process can be iterated, with the approximations being decomposed successively, so that one signal is broken down into many lower-resolution components. This is called the *multilevel wavelet decomposition*. Fig. 3 graphically represents this hierarchical decomposition:

Eq. (15) for the discrete wavelet expansion of a signal $s(t)$ can be employed in order to define the detail at level j . Let j be fixed and sum over the displacement k . A detail D_j is nothing more than the function

(definition of the detail at level j)

$$D_j(t) = \sum_{k \in Z} \text{DWT}(j, k)w_{j,k}(t). \quad (16)$$

$$\begin{aligned}
 S &= A_1 + D_1 \\
 &= A_2 + D_2 + D_1 \\
 &= A_3 + D_3 + D_2 + D_1
 \end{aligned}$$

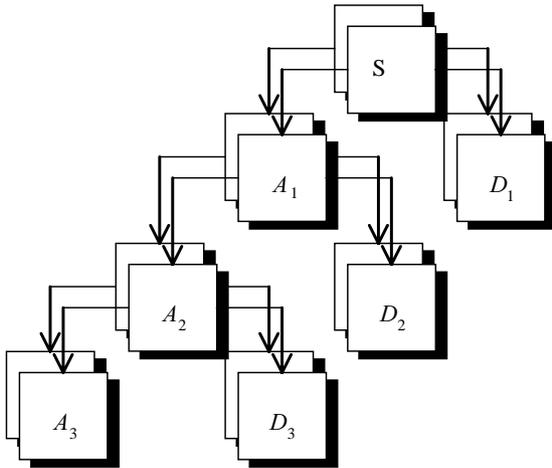


Fig. 3. Multilevel wavelet decomposition of a signal $s(t)$.

Now let the sum be over j . The signal is the sum of all the details:

(the signal is the sum of its details)

$$s(t) = \sum_{j \in \mathbb{Z}} D_j. \tag{17}$$

The details have just been defined. At reference level J , there are two types of detail signals. Those associated with indices $j \leq J$ correspond to the scales $a = 2^j \leq 2^J$ which are the fine details. The others, which correspond to $j > J$, are the coarser details. These latter signals are grouped into

(the approximation at level J) $A_J = \sum_{j > J} D_j$ (18)

which defines what is called an approximation of the signal $s(t)$. The details and an approximation at level J have thus been created. The equality

(Several decompositions) $s(t) = A_J + \sum_{j \leq J} D_j$ (19)

signifies that $s(t)$ is the sum of its approximation A_J and its fine details. From the previous formula, it is obvious that the approximation signals are related from level to level by

(link between A_{J-1} and A_J) $A_{J-1} = A_J + D_J$. (20)

2.6. The fast wavelet transform (FWT) algorithm

In 1988, Mallat (Misyty et al., 1996) proposed a fast wavelet decomposition and reconstruction algorithm. The Mallat algorithm for DWT is in fact a classical

scheme in the signal processing community, known as a two-channel subband coder using conjugate quadrature filters or quadrature mirror filters (QMF) (Masters, 1995). Mallat’s algorithm solves for the detail and approximation signals without finding the wavelet functions in Eq. (14). Mallat’s algorithm accomplishes for discrete wavelet analysis what Cooley’s and Tukey’s FFT algorithm accomplishes for Fourier analysis. Its steps, in short, are:

- The decomposition algorithm starts with the signal $s(t)$, and computes the values of the decomposed signals A_1 and D_1 , then those of A_2 and D_2 , and so on.
- The reconstruction algorithm called the *inverse discrete wavelet transform* (IDWT), starts with the signals A_J and D_J and uses them to compute A_{J-1} . Then from A_{J-1} and D_{J-1} it computes A_{J-2} and so on.
- Given a signal vector $s(t)$ of length n , the DWT proceeds in $\log_2 n$ steps at most. The first step starts from $s(t)$ and produces the following sets of coefficients: the approximation coefficients vector cA_1 and detail coefficients vector cD_1 . These coefficients are directly related to the $DWT(j, k)$ coefficients. They are obtained by convolving $s(t)$ with the low-pass filter LoF_D for the approximation, and with the high-pass filter HiF_D for the detail signals, followed by dyadic decimation (*down-sampling*). Fig. 4 shows a diagram of this procedure.

Let the length of each filter be equal to $2N$. If $n = \text{length}(s(t))$, the signals F and G are of length $n + 2N - 1$ and the coefficient vectors cA_1 and cD_1 are of length $\lfloor (n - 1)/2 \rfloor + N$.

The next step splits the approximation coefficients cA_1 into two parts, using the same scheme, producing cA_2 and cD_2 , and so on, as shown in Fig. 5.

Thus, the wavelet decomposition of the signal $s(t)$ analyzed at level j has the structure $[cA_j, cD_j, \dots, cD_1]$. This structure contains, for $j = 3$, the terminal nodes of the *wavelet decomposition tree* shown in Fig. 6.

In the reconstruction phase, starting from cA_j and cD_j , the IDWT reconstructs cA_{j-1} . Thus, it inverts the

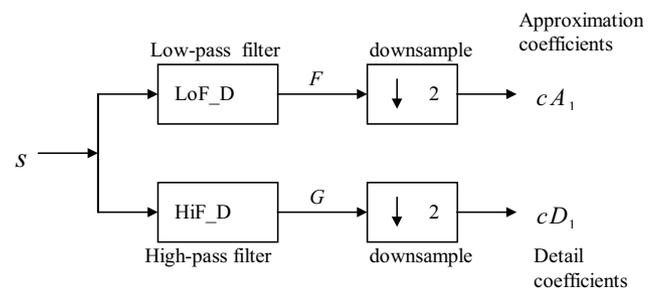


Fig. 4. First step of FWT algorithm.

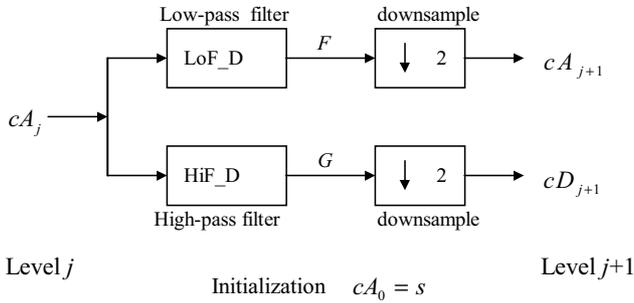


Fig. 5. One-dimensional DWT: decomposition step.

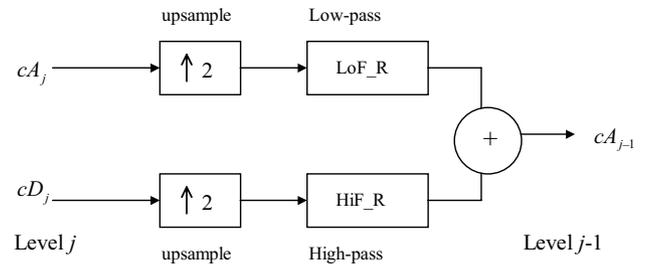


Fig. 7. One-dimensional IDWT: reconstruction step.

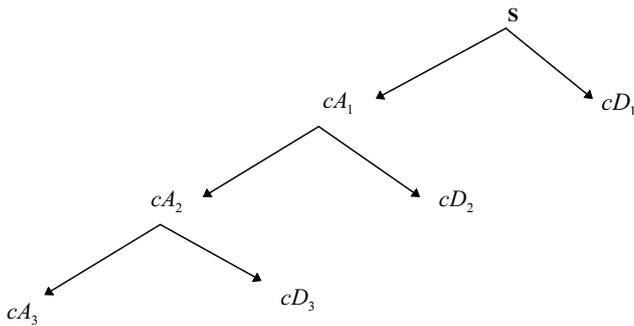


Fig. 6. Wavelet decomposition tree for $j = 3$.

decomposition step by inserting zeros between the samples of cA_j and cD_j and convolving the result with the corresponding reconstruction filters. This phase is shown in Fig. 7.

3. The use of wavelet analysis in pattern recognition

3.1. Fundamental concepts of pattern recognition

The problem of pattern recognition can be seen as one of classifying a group of objects on the basis of certain subjective similarity measures. Those objects classified into the same pattern class usually have some common properties. The classification requirements are subjective, since different classification occurs under different properties of the features (Banks, 1990; Tou and Gonzalez, 1974).

Given any particular pattern recognition problem, the first task is to choose a discretization method in order to obtain a *measurement vector* for each sample pattern. A major difficulty often arises when using these discretization methods, since the dimension of the *measurement space* is usually very large. It is therefore common practice to try to reduce this dimension by mapping the *measurement space* into a *feature space*, while retaining as many properties or *features* of the original samples as possible. This part of pattern recognition is called

feature selection (or preprocessing) and results in a set of samples from the *feature space* (Looney, 1997; Tate, 1996).

3.2. Feature extraction using wavelets

The wavelet transform may be used to represent efficiently the localized features of interest in a signal, which makes it an ideal tool for extraction of features and classification (Saito, 1994). It can be used as a filtering technique for removing the high-frequency components from the data, or as a method for representing shape information in a succinct way (Ogden, 1997). Alternatively, it has excellent data compression properties.

The use of the wavelet transform does not imply increase of the computational cost of the algorithm, as compared with the use of the Fourier transform. More specifically, for a signal of length N the fast wavelet transform has computational complexity of the order $O(N)$, whereas the fast Fourier transform has complexity of the order $O(N \log_2 N)$ (Galli et al., 1996; Ogden, 1997; Wilkinson and Cox, 1996).

In this paper, we consider signals that contain both transient and steady-state parts and combine features from classification from both parts. The analysis of the steady-state part has been well established. In our work we employ the Fourier transform on this part of the signal and extract features that relate to its stationary performance. More specifically, we consider the first eight odd harmonics of the steady-state vibration signal as potential features for classification. Alternatively, the transient part of the signal has not been studied for its potential in classification. In this paper, we also consider features for classification that are obtained from the wavelet coefficient vectors of the transient state of the signal. The wavelet transform algorithm operates by transforming the original signal vector (only its transient part) into a new one, which is filled sequentially with the wavelet coefficients of the different levels. The proposed algorithm for feature extraction from the transient part proceeds as follows:

We first compute the FWT of the transient state signal. The Daubechies wavelet function 4 (D4) with a resolution of five levels (levels 1, 2, 3, 4, 5) has been proven to be a good choice though this is not binding. The coefficients of all the components of fifth-level decomposition (that is, the fifth-level approximation and the first five levels of detail) are returned concatenated into one vector, \mathbf{C} . This vector is then split into the *detail wavelet coefficients* at individual levels, $\mathbf{cD1}$, $\mathbf{cD2}$, $\mathbf{cD3}$, $\mathbf{cD4}$, $\mathbf{cD5}$ and the *approximation wavelet coefficients*, $\mathbf{cA5}$ at level 5. These signals may exhibit some similarity or abrupt variations. In order to express signal similarity, the autocorrelation function is used, whereas a form of maximum deviation on smoothed signals to express rapid changes in the signal structure is exploited. These measures and the resulting features are presented next.

If $\mathbf{x}(n)$ is a sequence (vector) of length N , the sample autocorrelation function is calculated from

$$R_{xx}(l) = \sum_{n=i}^{N-|k|-1} \mathbf{x}(n)\mathbf{x}(n-l), \quad (21)$$

where $i = l$, $k = 0$ for $l \geq 0$, and $i = 0$, $k = l$ for $l < 0$. The index l is the time shift (or *lag*) parameter. We denote by \mathbf{AcDi} the autocorrelation of the signal \mathbf{cDi} .

The autocorrelation function may be viewed as a measure of similarity or *coherence*, between a signal $\mathbf{x}(n)$ and its shifted version (Ambardar, 1995). Clearly, under no shift, the two versions of the signal “match”, yielding the maximum for the autocorrelation function. But with increasing shift, it would be natural to accept the similarity and hence, the correlation between $\mathbf{x}(n)$ and its shifted version to decrease. As the shift approaches infinity, all traces of similarity vanish and the autocorrelation decays to zero. The autocorrelation function is symmetric about the origin where it attains its maximum value.

For smoothing rapid fluctuations a *signal-averaging filter* is used. A signal-averaging filter is also called a *smoothing filter* or *moving average filter* (Masters, 1995). This is done as follows:

Let $\mathbf{x}(n)$, $\mathbf{y}(n)$ be the input and output signals, respectively. For each data point $k \in [\xi, N - \xi]$ of $\mathbf{x}(n)$, the value

$$\mathbf{y}(k) = \frac{\sum_{l=k-m}^{k+m} |\mathbf{x}(l)|}{2m}$$

of $\mathbf{y}(n)$ is computed, where ξ is the starting data point of $\mathbf{x}(n)$, upon which the filter is operated, N the length of input signal, and $2m$ the window width of the input signal. If \mathbf{Si} denote the result of the filtering on the \mathbf{cDi} signals then as a measure of abrupt signal deviation, the quotients $\min |\mathbf{Si}| / \max |\mathbf{Si}|$ may be used.

Finally, sample variances and sample means of the \mathbf{cDi} and \mathbf{AcDi} signals computed by their usual

formulae, i.e.,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (22)$$

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (23)$$

may be added to the feature vector.

4. Case study: on-line quality control of washing machines

The aforementioned ideas have been successfully applied to the on-line quality control of washing machines. This application was carried out in the framework of the MEDEA project (MEDEA Final Report, 1999), a European Community funded project by the Standards, Measurement and Testing (SMT) action. The project was carried out by five European partners including AEA (Italy), MIT (Germany), CEA-LETI and CSO-Mesure (France), Universita degli Studi Ancona (Italy) and the Technical University of Crete. The project aimed at designing and building a prototype of an automatic system that could detect a range of mechanical defects in washing machines at the production line level. The defects of interest are reported in Table 1.

These five classes of defects (Z , B , P , M , H) are the most common according to a survey carried out in one of the major European Fairs, amongst all leading manufacturers.

Tests were carried out using two main sets of data. One was obtained in the laboratory of the Department of Mechanics of the University of Ancona, Italy (Paone et al., 1999), while the second was obtained at the prototype setup on the premises of AEA, Italy who is the exploiter of the final product. The two sets come from different types of washing machines. Originally, 11 points were chosen as candidates for possible reference points that could carry significant information in their vibration velocity signals regarding the health state of the washing machine. These points are shown in Fig. 8.

Table 1
Common defects present in washing machines (Domotecnica Appliances Fair, Cologne)

Defect class definitions	
Z	No defect
B	Electric motor clamping screws (released belt)
H	Releasing of the shock absorber
M	Use of different type of springs
P	Pulley (distorted)

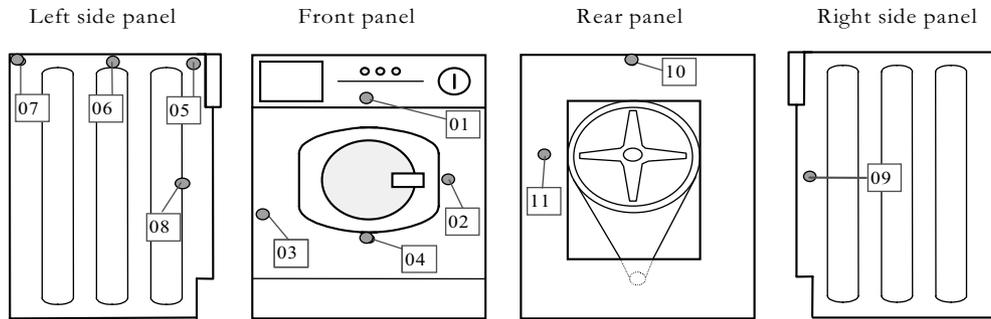


Fig. 8. The set of measurement points on the washing machine.

After preliminary investigation three points were chosen to test the proposed algorithms, namely 2, 8 and 9. This was done in order to reduce the computational complexity and the cost relevant to the measurement procedure. In both sets of data, two types of signals were measured: the vibration velocity at each measurement point and the rotational velocity of the washing machine's drum. This last measurement is used to separate the transient from the steady-state part of the vibration velocity signal.

In this study, three types of input features are extracted and compared in classification. More specifically, we consider:

- Fourier transform features from the steady-state (stationary) part of the signal, i.e. amplitudes at 8 odd harmonics of drum rotation frequency (Tselentis, 1998) (dimension of feature vector: 8 per point).
- Wavelet transform features from the transient (non-stationary) part of the signal, i.e. those described in Section 3.2 (dimension of feature vector: 10 per point).
- A combination of the above (dimension of feature vector: 18 per point).

Laser accelerometers are quite costly, and although part of the project involved the development of cheap sensors, the possibility of using fewer than three sensors was also investigated, since this could lead to a reduction of the total cost of the quality control system.

Classification performance was judged using a “leave-one-out” procedure, as follows:

for each machine k out of N
 train the classifier using $N - 1$ machines and leaving machine k out
 present machine k to the classifier
 end
 calculate machines correctly classified

The ratio $p = (\text{machines correctly classified}/N)$ then gives an indication of the relative merit of each combination.

The concept of pattern classification may be expressed in terms of the partition of the feature space (a mapping from feature space to *decision space*). Suppose that N features are to be measured from each input pattern. Each set of N features can be considered as a point in the N -dimensional feature space Ω_x . The problem of classification is to assign each possible vector or point in the feature space to a proper *pattern class*. This can be interpreted as a partition of the feature space into mutually exclusive regions, where each region corresponds to a particular pattern class (Barschdorff, 1991; Gose et al., 1996; Schalkoff, 1992).

The adopted classifier uses the following logic: assuming a normal distribution of pattern vectors in the feature space, the probability that a feature vector \mathbf{f} belongs to class $j (= 1, \dots, k)$ is given by (Barschdorff, 1991),

$$p(\mathbf{f}|j) = \frac{1}{((2\pi)^n \cdot \det[\mathbf{C}^j])^{1/2}} \times \exp[-0.5 \cdot (\mathbf{f} - \mathbf{m}^j)^T [\mathbf{C}^j]^{-1} (\mathbf{f} - \mathbf{m}^j)]$$

while the likelihood that \mathbf{f} originated from class j is given by

$$\ell_j = \frac{[p(\mathbf{f}|j) \cdot p(j)]}{p(\mathbf{f})}, \quad (24)$$

where $p(j)$ is the n -dimensional multivariable probability distribution of class j with mean \mathbf{m}_j and covariance \mathbf{C}_j and $p(\mathbf{f})$ denotes the prior probability that the feature vector belongs to class j . In this way, the fact that fault modes are less likely to occur is taken explicitly into account. If features are uncorrelated and normally distributed, Eq. (24) is easily calculated using

$$\ell_j = \ln[p(j)] - \frac{1}{2} \ln \det[\mathbf{C}^j] - \frac{1}{2} (\mathbf{f} - \mathbf{m}^j)^T [\mathbf{C}^j]^{-1} (\mathbf{f} - \mathbf{m}^j).$$

A fuzzy-like classifier can be obtained if the likelihoods are normalized,

$$L_k = \frac{1}{\sum_{j=1}^k \ell(j)} \begin{bmatrix} \ell(1) \\ \dots \\ \dots \\ \ell(k) \end{bmatrix}.$$

The i th element of this vector is the likelihood that the machine belongs to class i . Therefore, the decision is made that the machine belongs to the class that has the maximum likelihood

$$(\text{class index}) = \max_j(\ell(j)).$$

The first set of data, extracted from measurements taken in the laboratory environment, exhibit distinct boundaries for transient and steady-state parts of the rotational velocity as shown in Fig. 9.

As a result proper statistical algorithms could be used to separate the signal into its transient and steady-state part (Paone et al., 1999).

The relevant parts (transient and steady state) are then used into the feature extraction algorithms to produce Fourier and wavelet coefficients. The data are sampled at 2 kHz producing data files approximately containing 40 000 points for each sensor. The Fourier spectrum for the steady-state part has been restrained to frequencies that carry significant information, which is

up to 150 Hz. Furthermore, for the transient part, Figs. 10–12 show plots of approximation, detail wavelet coefficients, right-half parts of autocorrelation functions and moving average filtered detail coefficients of a typical washing machine vibration signal.

The total number of machines for the five points were 200 with class list

$$C_j = \{Z, B, P, M, H\}.$$

Prior probabilities for each class are calculated using the simple frequency formula (for the case where a machine type Z is left out for generalization):

$$p(j) = \left\{ \frac{39}{199} \quad \frac{40}{199} \quad \frac{40}{199} \quad \frac{40}{199} \quad \frac{40}{199} \right\}.$$

If only detection of defective (X) or non-defective machine (Z) is required, the class list is

$$C_j = \{Z, X\},$$

where

$$X = \{B, P, M, H\}$$

with prior probabilities (in the case where a type Z is left out)

$$p(j) = \left\{ \frac{39}{199} \quad \frac{160}{199} \right\}.$$

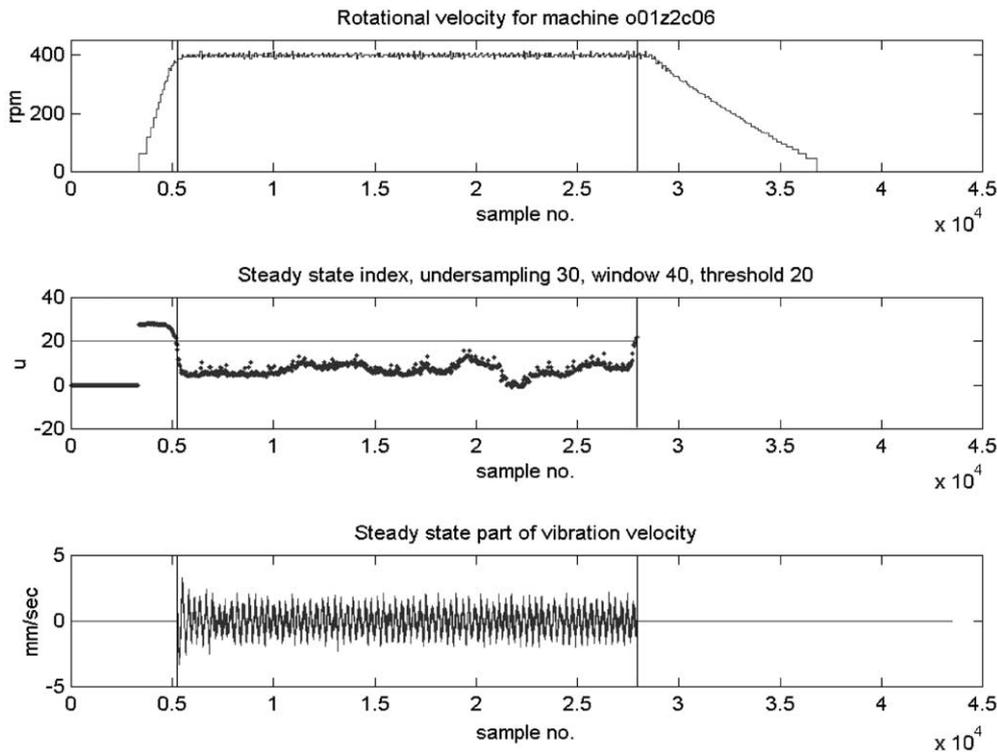


Fig. 9. Rotational velocity, separation index and stationary vibration velocity of a typical washing machine.

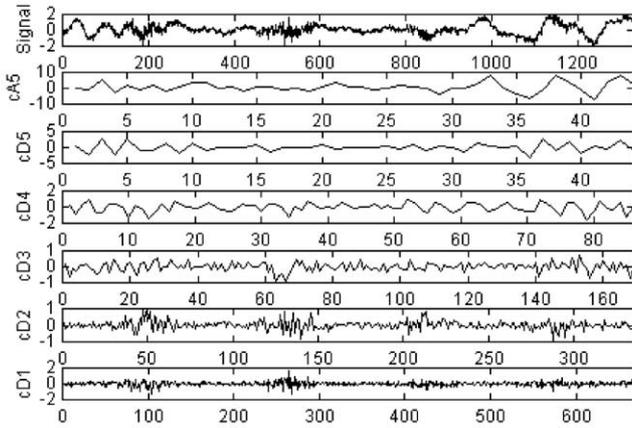


Fig. 10. Plots of transient vibration velocity signal S , vector of approximation $cA5$ and detail wavelet coefficients $cD1, cD2, cD3, cD4, cD5$ for a sample washing machine.

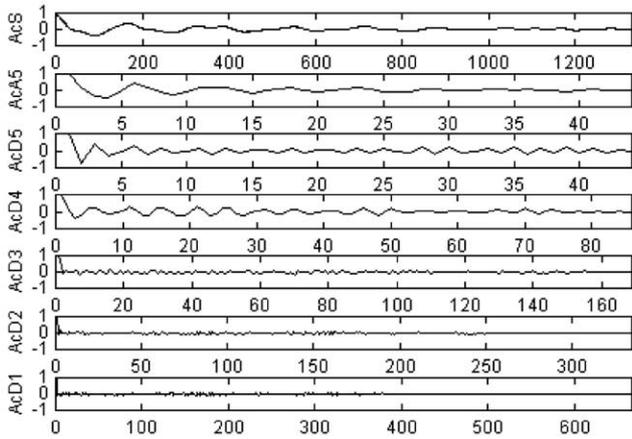


Fig. 11. Plots of right-half parts of the autocorrelation functions $AS, AcA5, AcD5, AcD4, AcD3, AcD2, AcD1$ for a sample washing machine.

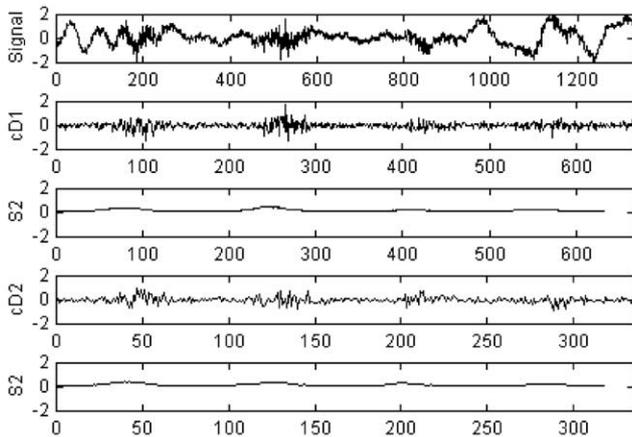


Fig. 12. Plots of $cD1, cD2$ signals with their outputs $S1, S2$ respectively after moving average filtering for a sample washing machine.

The results of the tests on these data are summarized in Tables 2–7.

The second set of data, extracted from measurements taken in the production line, do not exhibit distinct boundaries for transient and steady-state parts of the rotation velocity as shown in Fig. 13.

As a result proper statistical algorithms could not be used to separate the signal, which was therefore split using a heuristic, and rather arbitrary, method. The same feature extraction and classification procedures, as for the first set of data were applied to each part. The data are sampled at 10 kHz, producing data files approximately containing 160 000 points for each sensor. The Fourier spectrum for the steady-state part has again been restrained to frequencies up to 150 Hz. The frequency and wavelet domain features are extracted as before, through the proposed approach. The results are summarized in Tables 8–10.

Interpretation of the results yields the following general remarks. The best overall result, 99.5%, was obtained using two measurement points (2, 8) and Fourier transform features for discriminating amongst five classes (Table 2). This is an excellent performance, as it really made only one error in 200 samples. The best result using wavelet features, 85.5% was obtained using again the same two points but for discriminating among two classes (Table 5). Both these results were obtained using laboratory data, which is to be expected since most of the analysis was done using this set. The combination of Fourier and wavelet features increased the performance of the wavelet features (97%, Table 7) but still remained below the best score. The same behavior was observed on the production line data (Tables 8–10) but with worse results due to the aforementioned reasons.

The reduced performance of the features from the transient part of the signal is due to the high irregularity of the vibration signal at its transient phase. At this phase, the local features derived through the wavelet transform for each class of machines show large deviations in both their magnitudes and the locations they appear. Thus, features from different machine classes are inter-mixed, so that classes in this phase are not well separated in the feature space. To improve the performance of this type of features one needs bigger training sets of data for training the classifiers. In other words, we need to cover the feature space of each class more densely in the transient stage than in the steady-state operation. This is supported by the fact that the performance of the classifier used in this paper improves consistently with the size of the training set in the transient stage.

To summarize, wavelet transform features showed a promising performance when used as classification characteristics, but their use must always be judged

Table 2

Classification results using Fourier transform features on laboratory data (discrimination amongst five classes)

Class	Machines in sample	Point [8]		Points [2, 8]		Points [2, 8, 9]	
		Correctly classified	%	Correctly classified	%	Correctly classified	%
Z	40	31	0.775	40	1	40	1
B	40	40	1	40	1	40	1
P	40	33	0.825	40	1	38	0.950
M	40	38	0.950	40	1	39	0.975
H	40	40	1	39	0.975	39	0.975
Total	200	182	0.910	199	0.995	196	0.980

Table 3

Classification results using Fourier transform features on laboratory data (discrimination amongst two classes)

Class	Machines in sample	Point [8]		Points [2, 8]		Points [2, 8, 9]	
		Correctly classified	%	Correctly classified	%	Correctly classified	%
Z (healthy)	40	35	0.875	34	0.850	23	0.575
BPMH (faulty)	160	154	0.9625	160	1	160	1
Total	200	189	0.945	194	0.970	183	0.915

Table 4

Classification results using wavelet transform features on laboratory data (discrimination amongst five classes)

Class	Machines in sample	Point [8]		Points [2, 8]		Points [2, 8, 9]	
		Correctly classified	%	Correctly classified	%	Correctly classified	%
Z	40	11	0.275	31	0.775	33	0.825
B	40	33	0.825	27	0.675	20	0.500
P	40	15	0.375	21	0.525	17	0.425
M	40	17	0.425	23	0.575	18	0.450
H	40	26	0.6500	24	0.600	23	0.575
Total	200	102	0.510	126	0.630	111	0.555

Table 5

Classification results using wavelet transform features on laboratory data (discrimination amongst two classes)

Class	Machines in sample	Point [8]		Points [2, 8]		Points [2, 8, 9]	
		Correctly classified	%	Correctly classified	%	Correctly classified	%
Z (healthy)	40	7	0.175	19	0.475	9	0.225
BPMH (faulty)	160	153	0.9563	152	0.950	159	0.9938
Total	200	160	0.800	171	0.855	168	0.840

Table 6

Classification results using Fourier + wavelet transform features on laboratory data (discrimination amongst five classes)

Class	Machines in sample	Point [8]		Points [2, 8]		Points [2, 8, 9]
		Correctly classified	%	Correctly classified	%	
Z	40	31	0.775	40	1	Numerical instability problems
B	40	40	1	31	0.775	
P	40	33	0.825	27	0.675	
M	40	38	0.950	34	0.850	
H	40	40	1	27	0.675	
Total	200	182	0.910	159	0.795	

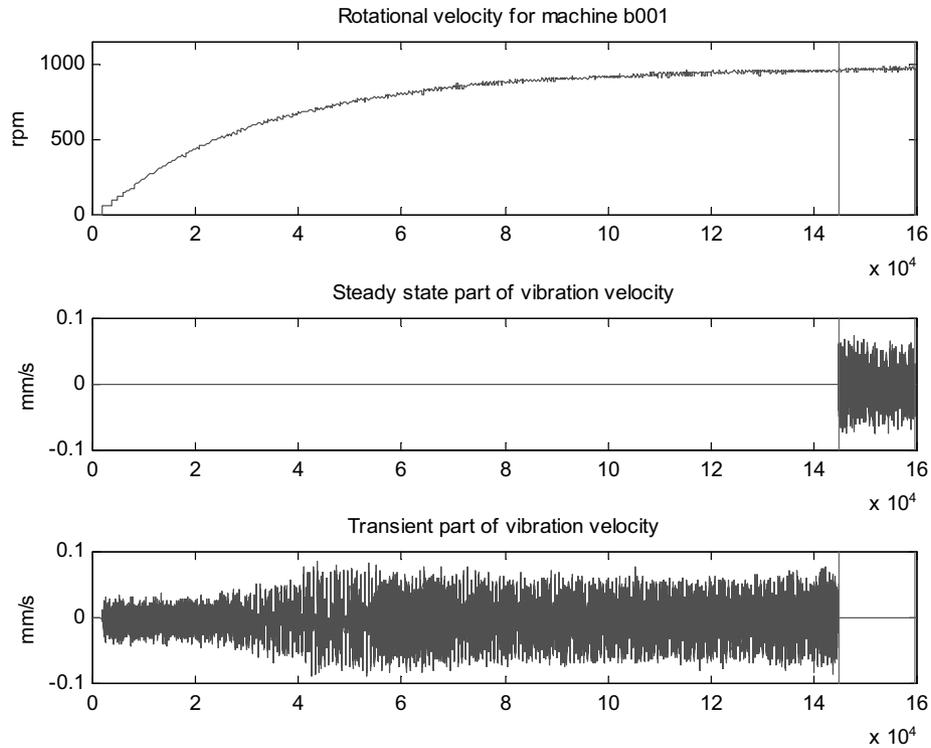


Fig. 13. Rotational velocity, steady state and transient part of vibration velocity of a typical washing machine from the production line.

Table 7
Classification results using Fourier + wavelet transform features on laboratory data (discrimination amongst two classes)

Class	Machines in sample	Point [2]		Points [2, 8]		Points [2, 8, 9]
		Correctly classified	%	Correctly classified	%	
Z (healthy)	40	34	0.850	6	0.150	Numerical instability problems
BPMH (faulty)	160	160	1	159	0.9938	
Total	200	194	0.970	165	0.825	

Table 8
Classification results using Fourier transform features on production line data (discrimination amongst two classes)

Class	Machines in sample	Points [2, 8, 9]	
		Correctly classified	%
Z (healthy)	52	41	0.7885
BPW (faulty)	61	51	0.8361
Total	113	92	0.8142

Table 10
Classification results using Fourier + wavelet transform features on production line data (discrimination amongst two classes)

Class	Machines in sample	Points [2, 8, 9]	
		Correctly classified	%
Z (healthy)	52	36	0.6923
BPW (faulty)	61	54	0.8852
Total	113	90	0.7965

Table 9
Classification results using wavelet transform features on production line data (discrimination amongst two classes)

Class	Machines in sample	Points [2, 8, 9]	
		Correctly classified	%
Z (healthy)	52	30	0.5769
BPW (faulty)	61	46	0.7541
Total	113	76	0.6726

against results obtained by classical Fourier transform features.

5. Conclusions

In this paper it is investigated that the applicability of features extracted from wavelet coefficients in the

problem of pattern recognition. These features can be used on their own or in conjunction with features extracted from the Fourier transform. The proposed method is tested on real data taken from a production line of washing machines. The aim is to classify produced machines according to their mechanical health state. Results show a promise in the use of wavelet-born features, but their performance is inferior to that of Fourier-based features. This could be due to the transient signal not carrying separation information or the inappropriateness of the proposed features. Further research will help in clarifying these issues.

References

- Ambardar, A., 1995. *Analog and Digital Signal Processing*. PWS Publishing Company, Boston.
- Barschdorff, D., 1991. Comparison of neural and classical decision algorithms. *Proceedings of the IFAC Symposium on Fault Detection, Supervision and Safety for Technical Process*. Baden-Baden, Germany, pp. 409–415.
- Banks, S., 1990. *Signal Processing-Image Processing and Pattern Recognition*. Prentice-Hall, United Kingdom.
- Galli, A., Heydt, G.T., Ribeiro, P.F., 1996. Exploring the power of wavelet analysis. *IEEE Computer Applications in Power* 9 (4), 37–41.
- Gose, E., Johansonbaugh, R., Jost, S., 1996. *Pattern Recognition and Image Analysis*. Prentice-Hall, New Jersey.
- Hayes, M., 1996. *Statistical Digital Signal Processing and Modeling*. Wiley, USA.
- Kim, Y., Rizzonni, G., Bahman, S., Wang, Y., 1995. Analysis and processing of shaft angular velocity signals in rotating machinery for diagnostic applications. *Proceedings International Conference on Acoustics, Speech and Signal Processing*, vol. 5, Detroit, MI, USA, 9–12 May, pp. 2971–2974.
- Krauss, T., Shure, L., Little, J., 1994. *Signal Processing Toolbox' for Use with MATLAB, User's Guide*. The Math Works, Inc., USA.
- Lee, N., Schwartz, S., 1995. Robust transient signal detection using the oversampled gabor representation. *IEEE Transactions on Signal Processing* 43 (6), 1498–1502.
- Looney, C., 1997. *Pattern Recognition using Neural Networks: Theory and Algorithms for Engineers and Scientists*. Oxford University Press, New York.
- Masters, T., 1995. *Neural, Novel and Hybrid Algorithms for Time Series Prediction*. Wiley, New York.
- MEDEA Final Report, 1999. Quality control for household appliances by on-line evaluation of mechanical defects. European Community Project: Standards, Measurement and Testing, SMC4-CT95-2016.
- Misity, M., Oppenheim, G., Poggi, J.-M., 1996. *Wavelet Toolbox for Use with Matlab, User's Guide*. The Math Works, Inc., USA.
- Newland, D., 1994a. Wavelet analysis of vibration, part I: theory. *Journal of Vibration and Acoustics* 116, 409–416.
- Newland, D., 1994b. Wavelet analysis of vibration, part II: wavelet maps. *Journal of Vibration and Acoustics* 116, 417–425.
- Ogden, R.T., 1997. *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser, USA.
- Paone, N., Scalise, P., Stavrakakis, G., Pouliezios, A., 1999. Fault detection for quality control of household appliances by non-invasive laser Doppler technique and likelihood classifier. *Measurement* 25, 237–247.
- Qian, S., Chen, D., 1996. *Joint Time-Frequency Analysis: Methods and Applications*. Prentice-Hall PTR, New Jersey.
- Saito, N., 1994. Local feature extraction and its applications using a library of bases. Ph.D. Thesis, Department of Statistics, Yale University, USA.
- Schalkoff, R., 1992. *Pattern Recognition: Statistical, Structural and Neural Approaches*. Wiley, New York.
- Sólnes, J., 1997. *Stochastic Process and Random Vibrations: Theory and Practice*. Wiley, New York.
- Strang, G., Nguyen, T., 1996. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, USA.
- Tamaki, K., Matsuoka, Y., Uno, M., Kawana, T., 1994. Wavelet transform based signal waveform identification: inspection of rotary compressor. *Proceedings of IECON '94—20th Annual IEEE Conference on Industrial Electronics*, vol. 3, Bologna, Italy, 5–9 September, pp. 1931–1936.
- Tate, A.-R., 1996. Pattern recognition analysis of in vivo magnetic resonance spectra. Ph.D. Thesis, University of Sussex, Brighton, UK.
- Theodoridis, S., Koutroubas, K., 1998. *Pattern Recognition*. Academic Press, San Diego, USA.
- Tou, J., Gonzalez, R., 1974. *Pattern Recognition Principles*. Addison Wesley Publishing Company, London.
- Tselentis, G., 1998. Fault diagnosis using vibration testing in an industrial washing machine production line. Ph.D. Thesis, Department of Production and Management Engineering, Technical University of Crete, Chania, Greece (in Greek).
- Vetterli, M., Kovacevic, J., 1995. *Wavelets and Subband Coding*. Prentice-Hall PTR, New Jersey.
- Wang, W.J., McFadden, P.D., 1994. Application of wavelets in non-stationary signal analysis for early gear damage diagnosis. *Proceedings of the International Conference on Vibration Engineering ICVE'94*, Beijing, China, pp. 615–620.
- Wilkinson, W., Cox, M., 1996. Discrete wavelet analysis of power system transients. *IEEE Transactions on Power Systems* 11 (4), 2038–2044.