Feature Extraction From Wavelet Coefficients for Pattern Recognition Tasks.

ARTICLE *in* IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE · JANUARY 1999

Impact Factor: 5.78 · DOI: 10.1109/34.745739 · Source: DBLP

2 AUTHORS, INCLUDING:

Sagar Kamarthi
Northeastern University
119 PUBLICATIONS **707** CITATIONS

SEE PROFILE

Feature Extraction From Wavelet Coefficients for Pattern Recognition Tasks

Stefan Pittner and Sagar V. Kamarthi

Abstract—In this paper, a new efficient feature extraction method based on the fast wavelet transform is presented. This paper especially deals with the assessment of process parameters or states in a given application using the features extracted from the wavelet coefficients of measured process signals. Since the parameter assessment using all wavelet coefficients will often turn out to be tedious or leads to inaccurate results, a preprocessing routine that computes robust features correlated to the process parameters of interest is highly desirable. The method presented divides the matrix of computed wavelet coefficients into clusters equal to rowvectors. The rows that represent important frequency ranges (for signal interpretation) have a larger number of clusters than the rows that represent less important frequency ranges. The features of a process signal are eventually calculated by the euclidean norms of the clusters. The effectiveness of this new method has been verified on a flank wear estimation problem in turning processes and on a problem of recognizing different kinds of lung sounds for diagnosis of pulmonary diseases.

Index Terms—Feature extraction, fast wavelet transform, signal interpretation.

1 Introduction

To analyze and interpret sensor signals monitoring a process in a certain application, reduction of the data size is an important preprocessing step in a nonstructural pattern recognition task [1]. But, only by using relevant features extracted from sensor signals is it possible to make inferences about the parameters or states of the process being monitored. In this paper, we use the term "process parameter" to refer to a continuous process variable, such as temperature, friction, or width of tool wear, and the term "process state" to refer to a process condition, such as good, out of control, normal, or unacceptable. Due to their time-frequency localization properties, discrete wavelet and wavelet packet transforms [2] have been proven to be appropriate starting points for the classification of the measured signals. They allow the extraction of richer problem-specific information from sensor signals than earlier methods for many practical applications. Based on these transforms, several methods have been developed to capture sensor signal features from which process parameters or states can be computed by using neural networks [3], [4], [5], [6] or certain statistical procedures [7], [8]. The overall computation scheme is depicted in Fig. 1. The input signals s consist of values uniformly sampled from a certain sensor. Unless stated otherwise, we assume throughout the paper that the signal interpretation routine is based on a neural network.

In Section 2, we present the mathematical background for the wavelet transform and introduce important methods for the final signal interpretation. Section 3 reviews existing methods for extracting features from wavelet coefficients. A new automatic feature extraction procedure is described in Section 4. This new method enables a large reduction of the wavelet transform data while retaining problem-specific information, which facilitates an efficient pattern recognition process. The effectiveness of the proposed method for flank wear estimation in turning processes and classification of lung sounds in diagnosis of pulmonary diseases is demonstrated in Section 5 and Section 6, respectively.

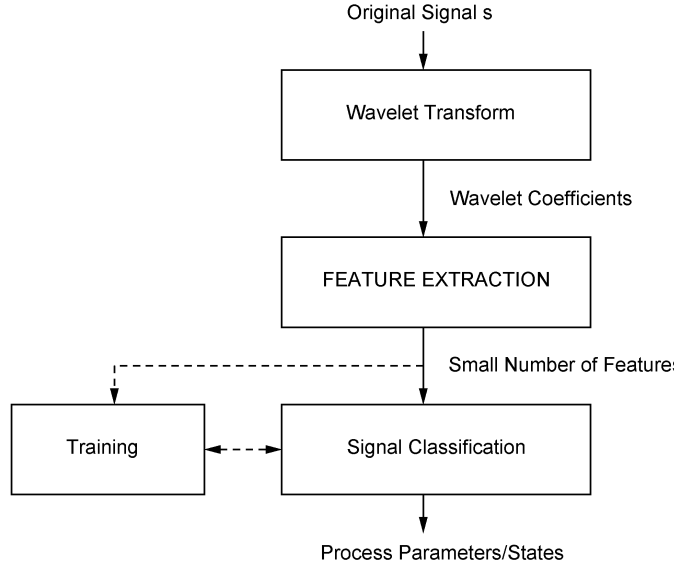

Fig. 1. Computation scheme for interpretation of process signals.

2 Theoretical Review

In this section, the theoretical foundations for the common wavelet-based feature extraction methods are presented.

2.1 The Fast Wavelet Transform

Due to desirable properties concerning approximation quality, redundancy, numerical stability, etc., the wavelet bases constructed by Daubechies [9], [10] became the foundation for the most popular techniques for signal analysis and representation in a wide range of applications. The parameters of the corresponding discrete wavelet decomposition are given by real-valued sequences $(h_0, h_1, \ldots, h_N)$ that solve the equations

$$\sum_{n=0}^{N} h_n = \sqrt{2} \tag{1}$$

$$\sum_{n=0}^{N} (-1)^n n^k h_n = 0 \qquad \text{for } k = 0, 1, \ldots, (N-1)/2 \tag{2}$$

$$\sum_{n=0}^{N-2k} h_n h_{n+2k} = 0 \qquad \text{for } k = 1, 2, \ldots, (N-1)/2 \tag{3}$$

for some odd natural number N; the number of solutions is equal to $2^{(N-1)/2}$.

The decomposition starts with a finite sequence $a_0 = (a_{0n})$ ($n = 0, 1, \ldots, L-1$) with a certain length L, which characterizes a measured signal s. This sequence a_0 is usually built from the sampling values of s, although better starting values are possible [11]. The sequences $a_1 = (a_{1n})$ and $d_1 = (d_{1n})$, defined by

$$a_{(m+1)n} = \sum_{k=2n}^{2n+N} h_{k-2n} a_{mk}, \tag{4}$$

$$\left(n = \tfrac{1-N}{2}, \tfrac{1-N}{2}+1, \ldots, \tfrac{L}{2}-1 \text{ for even } L \right.$$

$$\left. \text{and } n = \tfrac{1-N}{2}, \tfrac{1-N}{2}+1, \ldots, \tfrac{L-1}{2} \text{ for odd } L\right)$$

$$d_{(m+1)n} = \sum_{k=2n+1-N}^{2n+1} (-1)^k h_{2n+1-k} a_{mk} \tag{5}$$

• *The authors are with the Department of Mechanical, Industrial, and Manufacturing Engineering, 334 Snell Engineering Center, Northeastern University, Boston, MA 02115. E-mail: {spittner, sagar}@coe.neu.edu.*

Manuscript received 30 Apr. 1997; revised 19 Oct. 1998. Recommended for acceptance by R. Chellappa.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 108095.

$\left(n = 0, 1, \ldots, \frac{L+N-3}{2} \text{ for even } L \text{ and } n = 0, 1, \ldots, \frac{L+N-2}{2} \text{ for odd } L\right)$

with $m = 0$ have about half of the length of a_0. They represent a decomposition of the information contained in a_0, since a simple calculation which uses (1), (2), and (3) shows that

$$a_{0n} = \sum_{k=-\infty}^{\infty} \left[h_{n-2k} a_{1k} + (-1)^n h_{2k+1-n} d_{1k} \right] \text{ for every } n \in \mathbb{Z}, \quad (6)$$

i.e., the sequence a_0 can be precisely reconstructed from a_1 and d_1 [12]. Moreover, it can be proven that the energy of the sequence a_0 is divided among a_1 and d_1.

DEFINITION 1. *The energy $\|p\|^2$ of a finite sequence $p = (p_n)$ is defined by*

$$\|p\|^2 := \sum_{n=-\infty}^{\infty} p_n^2 . \quad (7)$$

The square root $\|p\|$ of the energy $\|p\|^2$ is also known as the Euclidean norm $\|p\|_2$ of p if p is regarded as a vector.

LEMMA 2. The energies of the sequences a_0, a_1, and d_1 satisfy the equation

$$\|a_0\|^2 = \|a_1\|^2 + \|d_1\|^2 . \quad (8)$$

Recursive applications of (4) and (5) lead to a decomposition of the sequence a_0 into a matrix of sequences d_1, d_2, …, d_M, a_M, as shown in Fig. 2. They contain the same amount of information as a_0 and have the property

$$\|a_0\|^2 = \|a_M\|^2 + \sum_{m=1}^{M} \|d_m\|^2 . \quad (9)$$

The computed coefficients in these sequences form the wavelet decomposition of the measured signal *s*. Each such linear mapping of a_0 is called a fast wavelet transform. Since $d_{m+1} = 0$ (up to edge values) for every constant sequence a_m, the (wavelet) coefficients in the sequences d_m can be interpreted as the details of the signal *s* at coarser and coarser resolutions as *m* is increased. It means that d_1 represents the highest frequency region, d_2 the next lower frequency region, and so on. Accordingly, the index *m* of a sequence $a_m = (a_{mn})$ or $d_m = (d_{mn})$ is called scale, and the coefficients a_{mn} and d_{mn} are called the approximation and the detail coefficients respectively. The total number of computed coefficients in the matrix shown in Fig. 2 is roughly equal to the length of the original sequence a_0.

The wavelet packet transform is a generalization of the procedure described above, which includes the possibility of further decomposing the computed sequences d_m in the same way as the decomposition of the sequences a_m. In contrast to the fast wavelet transform, it provides the possibility to zoom into any desired frequency range for further decomposition.

2.2 Computation of Process Parameters (or States)

The accuracy of the whole procedure for process parameter (or state) assessment depends on the feature extractor, the pattern recognition unit, and their compatibility. Neural networks are very powerful tools for pattern recognition. After training them, neural networks can be readily used for process parameter (or state) assessment without requiring any knowledge of the underlying system. However, to achieve faster learning in neural networks, it is, in general, necessary to preprocess their input information to eliminate irrelevant information from the inputs. This preprocessing is also supposed to contribute more accurate assessments of process parameters (or states) [13]. The most frequently used neural network architectures for wavelet-based process parameter assessment are multilayer perceptrons [14], adaptive resonance theory neural networks [15], and learning vector quantization networks [16].

Classification tasks are also typically performed using probabilistic models, where matching scores are computed to determine how close is the input pattern to reference exemplar patterns. Examples are provided by the minimum-distance pattern classifier [17] and the Bayes classifier [18]. Even for these statistical classifiers, it is extremely important that the input patterns are free from redundant or irrelevant information and correlate well with the process states.

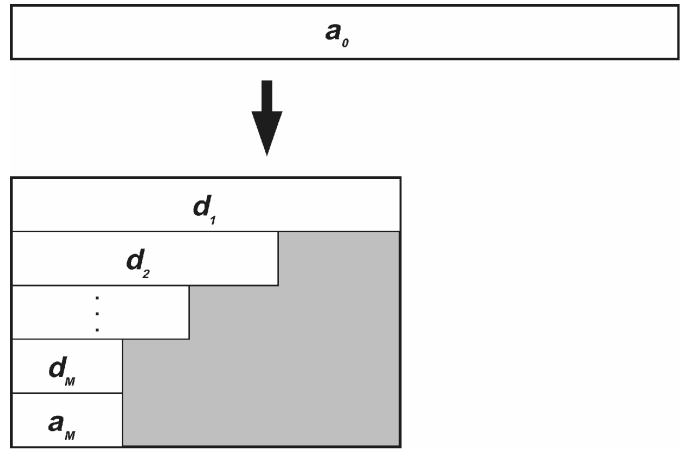

Fig. 2. Fast wavelet transform of a sequence a_0.

Due to the relatively large size of the output of the fast wavelet transforms or other spectral decompositions of signals *s*, the need for further dimensional reduction has been generally observed [19]. Informative features must be extracted from the coefficients computed with the fast wavelet transform of the process signals and used for training a classification or estimation neural network or a statistical procedure. For this purpose, a set of process signals, called training signals, together with the corresponding values of the process parameters (or states), are used for determining the weights of the estimator or the classifier with the features extracted from these training signals. In order to get correct assessment of process parameters (or states), the signals of the training set should reflect the entire process dynamics. When the neural network training is complete, features extracted from operational signals can be input to the neural network for estimation of process parameters or classification of process states.

3 CONVENTIONAL METHODS

In many existing methods, the energy values $\|d_m\|^2$ for $m = 1, 2, \ldots, M$ and $\|a_M\|^2$ are calculated based on all coefficients that belong to scales *m*. This helps to overcome the well-known problem posed by the fact that wavelet coefficients are not invariant to shifts in the process signals. In these methods, the energy values or the wavelet coefficients themselves for all or a selected number of scales *m* are chosen as the essential signal features. Here, the proper selection of a scale or a combination of scales is an important factor for obtaining good parameter assessment results. Two examples of such methods that select certain wavelet coefficients as signal features are described below.

Tansel et al. [20] developed a procedure to detect tool failure in an end milling process by isolating cutting force patterns and describing them with a small number of parameters. For every signal record, they calculated the sequence a_3 of approximation coefficients and presented it, up to three edge values, to a signal classification procedure provided by an ART2-type neural network. To verify the method, experiments were conducted at nine different cutting conditions. At each cutting condition, 10 experiments were conducted when tools are normal and nine experiments when the tools are broken. All of the nine cutting conditions, as well as the

tool conditions (normal or broken), were correctly classified by the ART2-type neural network using the wavelet features.

The automatic detection of epileptogenic waveforms in electroencephalograms was investigated by Kalayci and Özdamar [21]. They conducted several experiments, for each of which they fixed a combination of scales m between 1 and 5. During each experiment, they computed the eight central coefficients from the corresponding sequences d_m and used them as the features of the signal records. Taking those coefficients as inputs, a multilayer perceptron with one hidden layer gave classification accuracies up to 91 percent. The overall performance was observed to be better than when raw signals were input to neural networks.

This paper describes a new method which provides a refinement to the principle on which the most successful existing methods rely. Since good features should contain only problem-specific information of the input signals s and not much additional irrelevant information [22], the composition of such features is highly problem-dependent. Accordingly, in contrast to most of the prior methods, the feature extraction method proposed in this paper includes problem-specific parameters, which can be easily determined automatically for the problem at hand. In this proposed method, one or more features are extracted from all wavelet coefficients of each scale depending on its information content.

4 A New Cluster-Based Procedure

In this section, a novel approach of wavelet-based feature extraction for signal interpretation is introduced. As a preparation step to feature extraction, the wavelet coefficients are grouped into clusters in an unsupervised mode. These clusters are formed by using a set of representative signal records that pertain to a given application. In practical situations, this set can be the same as the set of signals used to train the estimator or classifier (e.g., a neural network). The following subsection describes the cluster formation method. The procedure proposed in this section divides the scheme of all computed wavelet coefficients into disjoint clusters $U_1, U_2, \ldots, U_c$ for each of which a single robust feature u_i ($i = 1, 2, \ldots, c$) can be computed. The so obtained feature vector $(u_1, u_2, \ldots, u_c)$ serves as an input pattern to a signal interpretation procedure such as a neural network.

4.1 Cluster Determination

The proposed method uses a set of K process signals called "representative signals" s that reflect the whole spectrum of possible process conditions to determine the boundaries of the clusters in an unsupervised mode. These cluster boundaries represent the problem-specific parameters of the feature extraction method. First, the complete fast wavelet decomposition is computed for all K representative signals s. The wavelet coefficients contained in the sequences $d_1, d_2, \ldots, d_M$ and a_M are arranged into a matrix for each individual signal s, as shown in Fig. 2. We denote each such matrix by $B = (b_{ij})$ ($i = 1, 2, \ldots, M + 1$; $j = 1, 2, \ldots, (L + N - 1)/2$ for even L; $j = 1, 2, \ldots, (L + N)/2$ for odd L). The shaded portion of B is filled with zero-valued elements.

The formation of the clusters from the wavelet coefficients is guided by the implicit assumption that the regions of B containing large-size wavelet coefficients allow better reconstruction and, primarily, better discriminability of the original signals than other regions of B. In addition to this heuristic concept, we make use of the following two results to derive the clustering procedure.

THEOREM 3 (Central Limit Theorem). *Let (Y_k) be a sequence of independent random variables such that $P(Y_k \in [\alpha, \beta]) = 1$ for every $k \in \mathbb{N}$ and some compact interval $[\alpha, \beta]$. Also, let*

$$\rho_n := \sqrt{\sum_{k=1}^{n} \mathrm{V}Y_k} \quad \text{and} \quad Z_n := \frac{\sum_{k=1}^{n} Y_k - \sum_{k=1}^{n} \mathrm{E}Y_k}{\rho_n} \tag{10}$$

for $n \in \mathbb{N}$, where $\mathrm{E}Y_k$ and $\mathrm{V}Y_k$ denote the expectation value and variance of a random variable Y_k, respectively. Then, $Z_n \xrightarrow{D} \mathcal{N}(0, 1)$ as $n \to \infty$ if and only if $\rho_n \to \infty$, where the symbol $\xrightarrow{D}$ means convergence in distribution.

A proof of this theorem was given by Sen and Singer [23].

LEMMA 4. *Let (Y_k) be a sequence of independent $\mathcal{N}(0, 1)$ distributed random variables, and $\gamma \geq e^2$ (e is the Euler number) be a constant. Then, for every $\epsilon > 0$, there exists a natural number $N(\epsilon)$ so that the expected value $\mathrm{E}Z_N$ of every random variable*

$$Z_N := \left| \left\{ Y_k : Y_k \geq \sqrt{2 \ln \frac{N}{\gamma}}, k = 1, 2, \ldots, N \right\} \right| \tag{11}$$

satisfies the inequalities

$$\frac{\gamma}{2\sqrt{\pi \ln N}} < \mathrm{E}Z_N < (1 + \epsilon) \frac{\gamma}{2\sqrt{\pi \ln N}} \tag{12}$$

for every natural number $N \geq N(\epsilon)$.

PROOF. The distribution function

$$F(x) := \mathrm{P}(Y_k \geq x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} \mathrm{d}t \tag{13}$$

for the standard normal distribution satisfies the inequalities

$$\frac{2}{\sqrt{x^2 + 4} + x} < \sqrt{2\pi} e^{\frac{x^2}{2}} F(x) < \frac{2}{\sqrt{x^2 + \frac{8}{\pi}} + x} \tag{14}$$

for every $x \in \mathbb{R}$, (see [24]). Taking $x = \sqrt{2 \ln(N/\gamma)}$ leads, after a few manipulations, to

$$\frac{\gamma}{2N\sqrt{\pi \ln N}} < F\left(\sqrt{2 \ln \frac{N}{\gamma}} \right) < (1 + \epsilon) \frac{\gamma}{2N\sqrt{\pi \ln N}} \tag{15}$$

for every sufficiently large value N. The result now follows from the fact that all independent random variables $\tilde{Z}_k$ ($k \in \mathbb{N}$) defined by

$$\tilde{Z}_k := \begin{cases} 1 & \text{for } Y_k \geq \sqrt{2 \ln \frac{N}{\gamma}} \\ 0 & \text{for } Y_k < \sqrt{2 \ln \frac{N}{\gamma}} \end{cases} \tag{16}$$

with the property

$$Z_N = \sum_{k=1}^{N} \tilde{Z}_k \tag{17}$$

have the same binomial distribution. □

Using the above results, the clusters can be determined from a representative set of K signals s. Let the matrix $\tilde{B} := (|b_{ij}|)$, constructed from the wavelet coefficients of the kth representative signal, be indexed by $\tilde{B}_k$ while $k = 1, 2, \ldots, K$. For any matrix A, we denote the sample mean and standard deviation of the elements of the matrix by $\mu(A)$ and $\sigma(A)$, respectively. Let I be the matrix of the same size as $\tilde{B}$ but containing only 1s as its elements. Let R be the operator which when applied to any matrix A, reduces the matrix by its last row.

It turns out that the elements of the matrix

TABLE 1
AN EXAMPLE OF THE CLUSTERING SCHEME OF COEFFICIENTS COMPUTED WITH THE FAST WAVELET TRANSFORM

Scale m	Wavelet Coefficients of a Signal s
1	$\{d_{1(0)}, d_{1(1)}, \ldots, d_{1(267)}\}$
2	$\{d_{2(0)}, d_{2(1)}, \ldots, d_{2(139)}\}$
3	$\{d_{3(0)}, d_{3(1)}, \ldots, d_{3(75)}\}$
4	$\{d_{4(0)}, d_{4(1)}, \ldots, d_{4(6)}\}\ \{d_{4(7)}, d_{4(8)}, \ldots, d_{4(43)}\}$
5	$\{d_{5(0)}, d_{5(1)}, \ldots, d_{5(27)}\}$
6	$\{d_{6(0)}, d_{6(1)}, \ldots, d_{6(10)}\}\ \{d_{6(11)}, d_{6(12)}, \ldots, d_{6(19)}\}$
7	$\{d_{7(0)}, d_{7(1)}, \ldots, d_{7(6)}\}\ \{d_{7(7)}, d_{7(8)}\}\ \{d_{7(9)}, d_{7(10)}\}\ \{d_{7(11)}, d_{7(12)}, \ldots, d_{7(15)}\}$
8	$\{d_{8(0)}, d_{8(1)}, \ldots, d_{8(5)}\}\ \{d_{8(6)}, d_{8(7)}\}\ \{d_{8(8)}\}\ \{d_{8(9)}, d_{8(10)}, \ldots, d_{8(13)}\}$
9	$\{d_{9(0)}, d_{9(1)}, \ldots, d_{9(6)}\}\ \{d_{9(7)}\}\ \{d_{9(8)}\}\ \{d_{9(9)}, d_{9(10)}, \ldots, d_{9(12)}\}$
9	$\{a_{9(0)}, a_{9(1)}, \ldots, a_{9(7)}, a_{9(8)}\}\ \{a_{9(9)}\}\ \{a_{9(10)}\}\ \{a_{9(11)}\}\ \{a_{9(12)}\}$

$$G = \left(g_{ij}\right) := \frac{1}{\sigma\left(R\left(\sum_{k=1}^{K} \tilde{B}_k\right)\right)}\left(\sum_{k=1}^{K} \tilde{B}_k - \mu\left(R\left(\sum_{k=1}^{K} \tilde{B}_k\right)\right) \cdot I\right) \quad (18)$$

reflect the size of the absolute values of a "typical wavelet decomposition" of the K selected signals, which are treated as the representative samples for all possible process signals. (The size of the matrix G equals the size of each matrix $\tilde{B}_k$.)

The elements at the same position in different matrices $\tilde{B}_k$ (k = 1, 2, …, K) can be considered as samples from independent random variables. This is due to the fact that they were computed for signal records measured at different time instances from the process being monitored. Consequently, Theorem 3 yields that the nonzero elements of $R(G)$ should be nearly $\mathcal{N}(0, 1)$ distributed. By applying a threshold of the form $T := \sqrt{2(\ln L - \ln \gamma)}$, with $\gamma \geq e^2$, to the elements of the matrix G, L being the number of computed detail coefficients, we get the binary matrix

$$G_b := (\Theta(g_{ij} - T)) \quad (19)$$

with the Heavyside function $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$. We expect $R(G_b)$ to contain at least $\gamma/2\sqrt{\pi \ln L}$ 1s by Lemma 4, although the independence of the random variables cannot be guaranteed in a general situation. The 1s in the matrix G_b occur at the same positions where the largest wavelet coefficients occur in the matrices B_k.

The clusters $U_1, U_2, \ldots, U_c$ are determined as rowvectors in G_b such that each cluster contains a 1 situated near the midpoint of the cluster. Clusters thus formed do not overlap across different scales. If a whole row of the matrix G_b contains no 1s, it is treated as a cluster by itself. Once the division of the matrix G_b into clusters is completed, the same pattern of clusters is superimposed on matrix B. With this pattern of clusters, regions of B containing important signal information, which are assumed to be characterized by larger size wavelet coefficients, are decomposed into a larger number of clusters than the regions of B containing less relevant information. See Table 1 for an example. In this table, for instance, scale 7 contains four clusters, while scale 6 contains two clusters. That means signal frequencies corresponding to scale 7 contain more useful information than frequencies corresponding to scale 6. The boundaries of the clusters $U_1, U_2, \ldots, U_c$ in a matrix B represent the application-specific grouping of wavelet coefficients and constitute the parameters of the feature extraction method.

4.2 Feature Extraction Method

From the procedure described in the previous subsection, clusters $U_1, U_2, \ldots, U_c$ are determined from a set of representative signals. The input feature vectors to a signal interpretation procedure can now be computed from measured process signals with a two-step procedure. In the first step, the fast wavelet transform is applied to a signal record s to obtain the matrix B of wavelet coefficients. This matrix B is divided into clusters $U_1, U_2, \ldots, U_c$ according to the same cluster boundaries as determined in the preceding subsection. Every rowvector that is formed by the elements of a cluster U_i in the matrix B is denoted by $\mathbf{r}_i$ for i = 1, 2, …, c. The second step consists of simply determining every component u_i of the feature vector $(u_1, u_2, \ldots, u_c)$ through the Euclidean norm

$$u_i := \|\mathbf{r}_i\|_2 = \sqrt{\sum_{v \in U_i} v^2} \quad (20)$$

of each such vector $\mathbf{r}_i$. This means that each feature u_i is determined as the square root of the energy of the wavelet coefficients in the corresponding cluster U_i. Consequently, the number c of features for a signal s is equal to the number of clusters determined with the method described in the previous subsection.

The physical meaning of each feature u_i is equal to that of a wavelet coefficient, namely it represents time and frequency information of the regarded signal s. A single feature especially describes a certain frequency range (a scale m), which is equal to that described by the wavelet coefficients underlying this feature. In contrast to that, each feature represents a connected time range of the signal s, whose resolution is as large as that of all underlying wavelet coefficients together. This time resolution can be as high as that of a single wavelet coefficient for information-rich scales with as many clusters as the number of wavelet coefficients. On the other hand, the time resolution can be so low that a feature describes the time range of the whole signal s for scales that consist of only one cluster.

The result

$$\|a_0\|^2 = \sum_{i=1}^{c} u_i^2, \quad (21)$$

which follows from (9), shows that the so constructed feature vectors are robust to noise in the corresponding signals s:

THEOREM 5. *Assume that s is a discrete signal with finite length and the fast wavelet decomposition is always started with the original signal, i.e., $a_0 = s$. For any signal $\tilde{s}$ that results by contaminating s with additive noise represented by a discrete signal w with the same length as s, i.e., $\tilde{s} = s + w$, and any possible cluster decomposition of the wavelet transform scheme, the feature vectors*

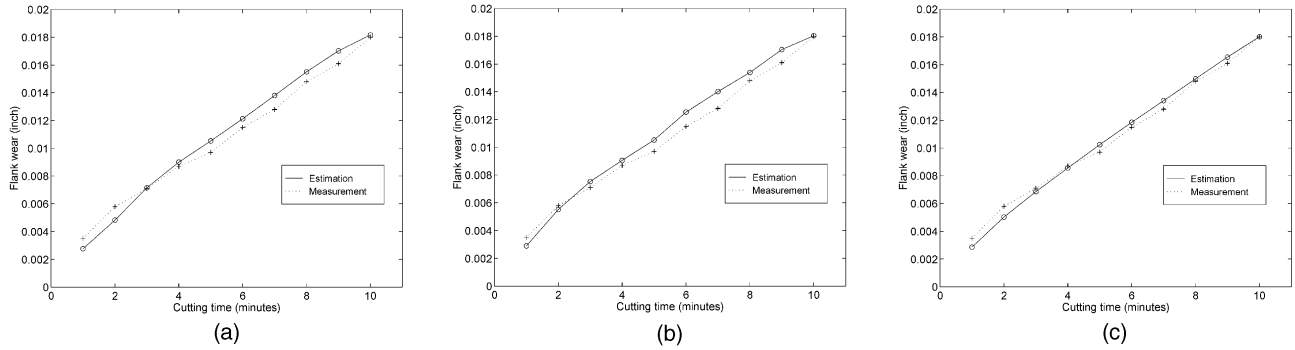

Fig. 3. Samples of flank wear estimates provided with the proposed feature extraction method from (a) force signals, (b) vibration signals, and (c) acoustic emission signals.

$(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_c)$ and $(u_1, u_2, \ldots, u_c)$ corresponding to $\tilde{s}$ and s, respectively, satisfy the inequality

$$\left\|(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_c) - (u_1, u_2, \ldots, u_c)\right\|_2 \le \|w\|_2. \quad (22)$$

PROOF. Let $\{U_1, U_2, \ldots, U_c\}$ be an arbitrary clustering of the scheme of wavelet coefficients for the signals s and $\tilde{s}$ as described in Section 4.1 and consider the fast wavelet transform of the signals s, $\tilde{s}$, and w. With the notation $\mathbf{r}_i$, $\tilde{\mathbf{r}}_i$, and $\mathbf{w}_i$ (i = 1, 2, …, c) for the rowvectors which are formed by the elements of a cluster U_i for s, $\tilde{s}$, and w, respectively, it follows from the linearity of the fast wavelet transform that

$$\tilde{\mathbf{r}}_i = \mathbf{r}_i + \mathbf{w}_i \quad \text{for } i = 1, 2, \ldots, c. \quad (23)$$

Consequently, the difference of corresponding features of the signals $\tilde{s}$ and s can be expressed by

$$\left|\tilde{u}_i - u_i\right| = \left|\left\|\mathbf{r}_i + \mathbf{w}_i\right\|_2 - \left\|\mathbf{r}_i\right\|_2\right| \quad (24)$$

for i = 1, 2, …, c. A standard inequality for norms further gives

$$\left|\tilde{u}_i - u_i\right| \le \left\|(\mathbf{r}_i + \mathbf{w}_i) - \mathbf{r}_i\right\|_2 = \left\|\mathbf{w}_i\right\|_2, \quad (25)$$

and, finally, with (21)

$$\left\|(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_c) - (u_1, u_2, \ldots, u_c)\right\|_2 = \sqrt{\sum_{i=1}^{c} (\tilde{u}_i - u_i)^2}$$
$$\le \sqrt{\sum_{i=1}^{c} \left\|w_i\right\|_2^2} = \|w\|_2. \quad (26)$$

□

The effectiveness of the proposed feature extraction method is demonstrated on two different applications in the following sections. Further examples and comparisons of the performance of the proposed method with that of other feature extraction procedures can be found elsewhere [11], [25], [26], [27].

5 Flank Wear Estimation

We used the procedure described in the previous sections to process data from force, vibration, and acoustic emission sensors to find the input patterns to a simple recurrent neural network for continuous flank wear estimation in turning processes [11], [28]. Force, vibration, and acoustic emission signals were sampled at 3 kHz, 26 kHz, and 1 MHz frequency, respectively. Sample records consisted of 512 points measured once in every minute as the cutting tool removes material from the workpiece. Experiments were divided into two different sets, Set 1 and Set 2. The data from Set 1 experiments were used to determine the boundaries of the clusters for the feature extraction as described in Section 4. The number of clusters formed was around 25 for different types of signals. With this scheme, features were extracted from the Set 1 and Set 2 data. The input patterns created from Set 1 data were used for training the neural network. The performance and accuracy of flank wear estimation was studied using the input patterns created from Set 2 data. Set 1 had 168 input patterns and Set 2 had 95 input patterns. Sample results are depicted in Fig. 3. The mean error and root-mean-square error for all three sensor types fall in the range from 0.0004 to 0.0009 inches and from 0.0011 to 0.0014 inches, respectively. This considerably low estimation error compared with other methods demonstrates the effectiveness of the feature extraction method proposed in this paper.

6 Classification of Lung Sounds

The automatic classification of different lung sound signals is the second application of the proposed feature extraction method. Each signal record consisted of 1,800 data points sampled at 2 kHz. Each one was to be assigned to one of the following five classes:

- inspiratory phase of bronchial sound,
- expiratory phase of bronchial sound,
- sound with crackles,
- sound with rhonhi, and
- sound with wheezes.

The first two classes represent normal lung sounds. The other three classes characterize abnormal lung sounds, namely short explosive sounds (crackles), continuous low-pitched sounds (rhonhi), and continuous high-pitched sounds (wheezes).

About 18 signals from each class were used to determine the input feature vectors. The size of the feature vectors was observed to be 26. Thereupon, a standard backpropagation neural network was trained based on the corresponding $18 \times 5 = 90$ feature vectors with 10,000 epochs for signal classification. For this purpose, the input feature vectors were randomly ordered. The network structure consisted of 26 hidden nodes and five output nodes representing the five different classes. The network was trained such that an output node responds with the value 0.95 or 0.05, depending on whether or not the feature vector presented belongs to the class represented by this output node. In this experiment, 100 percent correct classification was obtained for a different set of 40 test signals almost evenly distributed among the five signal classes. It clearly indicates the effectiveness and the reliability of the proposed approach for extracting features from sensor signals for the purpose of pattern classification or parameter estimation.

7 Conclusions

The study investigates a preprocessing method for reducing the size of input signals presented to neural networks for increasing the estimation or classification accuracy by retaining most of the intrinsic

information content of the measured signals. The stated results show that the proposed method can efficiently extract important features related to gradually increasing flank wear from wavelet expansions of sensor signals from a turning process. The proposed preprocessing has proven to be also readily applicable to other signal interpretation applications such as lung sound classification to reduce input record size and / or improve the performance.

REFERENCES

[1] R.J. Schalkoff, *Pattern Recognition: Statistical, Structural and Neural Approaches.* New York: John Wiley and Sons, 1992.

[2] C.K. Chui, *An Introduction to Wavelets.* Boston: Academic Press, 1992.

[3] A. Laine and J. Fan, "Texture Classification by Wavelet Packet Signatures," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 11, pp. 1,186–1,191, Nov. 1993.

[4] S. Livens, P. Scheunders, G.V. de Wouwer, D.V. Dyck, H. Smets, J. Winkelmans, and W. Bogaerts, "Classification of Corrosion Images by Wavelet Signatures and LVQ Networks," *Computer Analysis of Images and Patterns,* V. Hlavác and R. Sára, eds., pp. 538-543. Berlin: Springer-Verlag, 1995.

[5] J.-M. Nicolas, A. Lemer, and D. Legitimus, "Identification automatique de bruits impulsifs en acoustique sous-marine par résaux multicouches," *Revue Générale de l'Electricité,* vol. 5, pp. 43-48, May 1990.

[6] I.N. Tansel, C. Mekdeci, O. Rodriguez, and B. Uragun, "Monitoring Drill Conditions with Wavelet Based Encoding and Neural Networks," *Int'l J. Machine Tools and Manufacture,* vol. 33, pp. 559-575, Aug. 1993.

[7] A. Petrosian, R. Homan, S. Pemmaraju, and S. Mitra, "Wavelet Based Texture Analysis of EEG Signal for Prediction of Epileptic Seizure," *Wavelet Applications in Signal and Image Processing III,* SPIE, vol. 2,569, pp. 189–194, 1995.

[8] Y. Wu and R. Du, "Feature Extraction and Assessment Using Wavelet Packets for Monitoring of Machining Processes," *Mechanical Systems and Signal Processing,* vol. 10, pp. 29-53, Jan. 1996.

[9] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," *Comm. Pure and Applied Mathematics,* vol. 41, pp. 909-996, Oct. 1988.

[10] I. Daubechies, *Ten Lectures on Wavelets.* Philadelphia: SIAM Press, 1992.

[11] S.V. Kamarthi and S. Pittner, "Fast Fourier and Wavelet Transform for Flank Wear Estimation—A Comparison," *Mechanical Systems and Signal Processing,* vol. 11, pp. 791-809, Nov. 1997.

[12] S. Pittner, "Dyadic Orthonormal Wavelet Bases and Related Possibilities for Optimal Analysis and Representation Of One-Dimensional Signals," PhD thesis, Vienna Univ. of Technology, 1994.

[13] C.M. Bishop, *Neural Networks for Pattern Recognition.* Oxford: Clarendon Press, 1995.

[14] R.P. Lippmann, "An Introduction to Computing with Neural Nets," *IEEE Acoustics, Speech, and Signal Processing Magazine,* vol. 4, pp. 4-22, Apr. 1987.

[15] A.K. Jain, J. Mao, and K.M. Mohiuddin, "Artificial Neural Networks: A Tutorial," *Computer,* vol. 29, pp. 31-44, Mar. 1996.

[16] B.D. Ripley, *Pattern Recognition and Neural Networks.* Cambridge: Cambridge Univ. Press, 1996.

[17] J.T. Tou and R.C. Gonzalez, *Pattern Recognition Principles,* fourth ed. Reading, Mass.: Addison-Wesley, 1981.

[18] K.S. Fu, *Digital Pattern Recognition.* Berlin: Springer-Verlag, 1976.

[19] Y. Mallet, D. Coomans, J. Kautsky, and O. De Vel, "Classification Using Adaptive Wavelets for Feature Extraction," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, pp. 1,058–1,066, Oct. 1997.

[20] I.N. Tansel, C. Mekdeci, and C. McLaughlin, "Detection of Tool Failure in End Milling with Wavelet Transformations and Neural Networks (WT-NN)," *Int'l J. Machine Tools and Manufacture,* vol. 35, pp. 1,137-1,147, Aug. 1995.

[21] T. Kalayci and Ö. Özdamar, "Wavelet Preprocessing for Automated Neural Network Detection of EEG Spikes," *IEEE Eng. in Medicine and Biology,* vol. 14, pp. 160-166, Mar. / Apr. 1995.

[22] C.G. Looney, *Pattern Recognition Using Neural Networks: Theory and Algorithms for Engineers and Scientists.* New York: Oxford Univ. Press, 1997.

[23] P.K. Sen and J.M. Singer, *Large Sample Methods in Statistics—An Introduction with Applications.* New York: Chapman & Hall, 1993.

[24] D.S. Mitrinovic, *Analytic Inequalities.* Berlin: Springer-Verlag, 1970.

[25] S. Pittner, S.V. Kamarthi, and Q. Gao, "Wavelet Networks for Sensor Signal Classification in Flank Wear Assessment," *J. Intelligent Manufacturing,* vol. 9, pp. 315-322, Aug. 1998.

[26] S. Pittner and S.V. Kamarthi, "Image Features from One-Dimensional Wavelet Transforms for Defect Detection," *Intelligent Eng. Systems Through Artificial Neural Networks,* vol. 7, C.H. Dagli, M. Akay, O. Ersoy, B.R. Fernández, and A. Smith, eds., pp. 561-566. New York: Am. Soc. Mechanical Engineers, 1997.

[27] S. Pittner, A. Hoque, and S.V. Kamarthi, "Image Features for Defect Detection—Comparison of One- and Two-Dimensional Wavelet Transforms," *Intelligent Eng. Systems Through Artificial Neural Networks,* vol. 8, C.H. Dagli, M. Akay, A.L. Buczak, O. Ersoy, and B.R. Fernández, eds., pp. 479-484. New York: Am. Soc. Mechanical Engineers, 1998.

[28] S.V. Kamarthi, S.R.T. Kumara, and P.H. Cohen, "Flank Wear Estimation in Turning Through Wavelet Representation of Acoustic Emission Signals," *J. Eng. for Industry,* to appear.