Detection and Classification of Harmonic Transients by Using Trigonometric Smooth Wavelet-Packets

Erkennung und Klassifizierung von harmonischen Transienten mit trigonometrischen glatten Wavelet-Paketen

Paolo Mercorelli, Peter Terwiesch and Domenico Prattichizzo

The proposed contribution describes a method for the detection and classification of harmonic transients, as they occur in electrical networks. The method uses, as bases, a special class of wavelet-packets, i. e. smooth biorthogonal sine and cosine-packets in order to find the best "Shannon-basis". The algorithm is based on the minimization of Shannon's function for the decomposed wanted signal over the biorthogonal analyzing functions. The proposed method is validated by using a real application, signals from rail systems. One shows the fastness and the robustness of the algorithm and thus its real time applicability. In particular, one detects and classifies inrush currents.

Der vorliegende Beitrag beschreibt ein Verfahren zur Erkennung und Klassifizierung von harmonischen Transienten, wie sie unter anderem in elektrischen Netzen auftreten. Das Verfahren verwendet als Basis eine spezielle Klasse von Wavelet-Paketen, nämlich glatte lokalisierte biorthogonale Sinus-und Kosinus-Pakete, um für ein klassifizierendes Signal die beste „Shannon-Basis" zu finden. Dieser Signalzerlegung liegt die Minimierung der Shannon'schen Entropiefunktion zugrunde, sie hängt also adaptiv vom zu analysierenden Signal ab. Die Praxistauglichkeit des sehr schnellen echtzeitfähigen Verfahrens wird anhand eines praktischen Beispiels aus der Bahntechnik illustriert, bei dem für die Schutzabschaltung zwischenbetriebstypische Spannungsänderungen aufgrund der Einfahrt in einen neutralen Netzabschnitt und sicherheitskritische Spannungsänderungen aufgrund interner Fehler unterschieden werden müssen.

Schlagwörter: Klassifizierung, harmonische Transienten, Wavelet-Paketen

Keywords: Classification, harmonic transients, wavelet-packets

1 Introduction and Motivations

The goal of this paper is to introduce an algorithm for detection and classification of time-frequency phenomena, specifically of transient harmonics as they occur in electrical power systems [22; 25]. The signals under consideration are typically of a quasi-harmonic character and change amplitude and phase only slowly, making them comparatively narrow-banded. Recent work in this direction [22; 25] has indicated wavelets as a promising approach for off-line analysis and classification. However, a fast algorithm that can reliably distinguish different transients in real time based on a minimum number of data points remains to be found.

In an electrical railway vehicle, such an algorithm could be used to distinguish a harmless transformer inrush current, which is caused by sudden changes in line voltage, e.g., after passing a neutral section, from other, potentially hazardous, transients that would require an immediate shutdown of the vehicle.

Artificial Neural Network have shown success in solving classification problems. However, in designing a classification system several choices are required. First, a decision needs to be made on the particular neural network model and training method. In other words one has to choose: the activation function, the number of layers and the kind of algorithm. In fact, different sets of features are extracted using different activation functions and extraction methods. Finally, a choice on method of validation which gives some bound of the classification error rate. Unfortunately, only loose guidelines exist which govern any of these choices. Thus, decisions which influence the classification success percentage of the classifier base often on little more than intuition or even random chance. However, classification problems can be made to look like nonparametric regression if the outputs of the estimated function are interpreted as being proportional to the probability that the input belongs to the corresponding class.

Recently, the theory of wavelet packets has emerged as an alternate time-frequency analysis tool to the Fourier transform. Wavelets have been applied to a variety of problems, mostly in data compression and noise reduction, later in nonparametric modelling estimation and very few in feature extraction and classification. It is reasonable to investigate the application of the theory of wavelets to the problem of feature extraction.

Furthermore, in literature, as [22; 25], the wavelet field is indicated as the direction where to investigate in order to develop a novel method for detecting and classifying disturbances and transients regarding the power quality problem. Our specific goal is to build a classification method for narrow band signals, i. e. signals with slowly varying amplitude and phase even though the algorithm presented is totally general and can be applied for every signal but its effectiveness is for a particular class of signals or harmonic detection.

Progress in this direction is marked in [21] which proposes an efficient algorithm in order to model strong nonlinear systems, for instance in [26] one detects and classifies transients. In on-line detection of harmonic features interesting contributions have been presented in [20], where solutions to the problem of detecting dominate frequency vibration in pantograph system are proposed.

This particular class of signals or harmonic problems are often present in electrical power system, for instance, signals or disturbances with localized band around the multiple of the fundamental frequency. Transformer inrush current is a typical application and an important problem which occurs due to the saturation of non linearity of the transformer. For example, when the transformer of an electrical vehicle is connected to the overhead line because of the discontinuity of the magnetic flux, it arises an over tension which generates a consequent over current: the saturation occurs. In general, behind these scopes there are several interesting methodological approaches which can be *ad hoc* consequently investigated and developed.

The present paper proposes an algorithm for classifying the signals with sparse training data combining techniques in regression analysis and backpropagation procedures. The structural biorthogonality of the bases guarantees robustness and efficient numerical calculations. The procedure uses nonorthogonal bases (frames). By relaxing the orthogonality much more freedom on the choice of the wavelets is gained, though suitable only for problems with small dimensions. In order to consider and to use the nonorthogonality of the frames, which generates an interaction between the elements of the bases, the algorithm considers to every step again all the elements of the bases previously selected, without any elimination, [27]. The procedure is similar but simpler than the *projection pursuit regression* in [19] or the *stepwise selection by orthogonalization* in [27]. The paper is organized as follows. In Sect. 2 the smooth trigonometric wavelet packets are briefly discussed. In Sect. 3 and in Sect. 4 one defines the analytic problem and several basic issues are briefly discussed. Section 5 is devoted to the mathematical description of the proposed algorithm and in Sect. 6 a real application is given.

2 Wavelet Frames and Biorthogonal Smooth Trigonometric Wavelet Packets

Wavelet transform and wavelet series are becoming popular in signal processing and numerical analysis. Loosely speaking, a function $f(t)$ can be decomposed into

$$f(t) = \sum_j \sum_n w_{j,n} \psi_{j,n}(t) \tag{1}$$

where the $\psi_{j,n}(t)$ are the wavelet functions, normally obtained by dilating and translating a mother function $\psi(t)$, the index j and n denote the dilation and translation respectively and $w_{j,n}$ is the weight coefficient for $\psi_{j,n}(t)$. The most popular algorithms are related to the orthonormal wavelet bases ([4] or [8]) characterized by fast and elegant algorithms. Besides these, there are, less used ones, the *wavelet frames*, which have been introduced the first time in [11] and for which the computations of the coefficients are more complicated. But still, they have certain advantages. As wavelet frames consist of nonorthogonal wavelet families, they are *redundant bases*. To be more formal:

Definition 1. *A family of functions* $\{\psi_{j,n}(t); (j,n) \in Z, t \in \Re\}$ *in a Hilbert space* $\mathcal{H}$ *is called a frame of* $\mathcal{H}$ *if for every element* $f(t) \in \mathcal{H}$ *there are two positive constants* **A** *and* **B** *such that:*

$$\mathbf{A}\| f(t)\|^2 \le \sum_{j,n} \|\langle f(t), \psi_{j,n}(t)\rangle\|^2 \le \mathbf{B}\| f(t)\|^2 . \tag{2}$$

Where $\langle \cdot, \cdot \rangle$ indicates the inner product and $\| \cdot \|$ the norm.

It is already known that under the frame condition, $f(t)$ can be recovered from $\langle f(t), \psi_{j,n}(t)\rangle$ with some iterative procedures [8; 17]. Hence, if the set of bases constitutes a frame, the reconstruction of the function $f(t)$ in (1) is ensured. For practical implementation, infinite wavelet frames must be truncated into finite sets.

By relaxing the orthogonality, much more freedom on the choice of the wavelet functions is gained and elasticity for

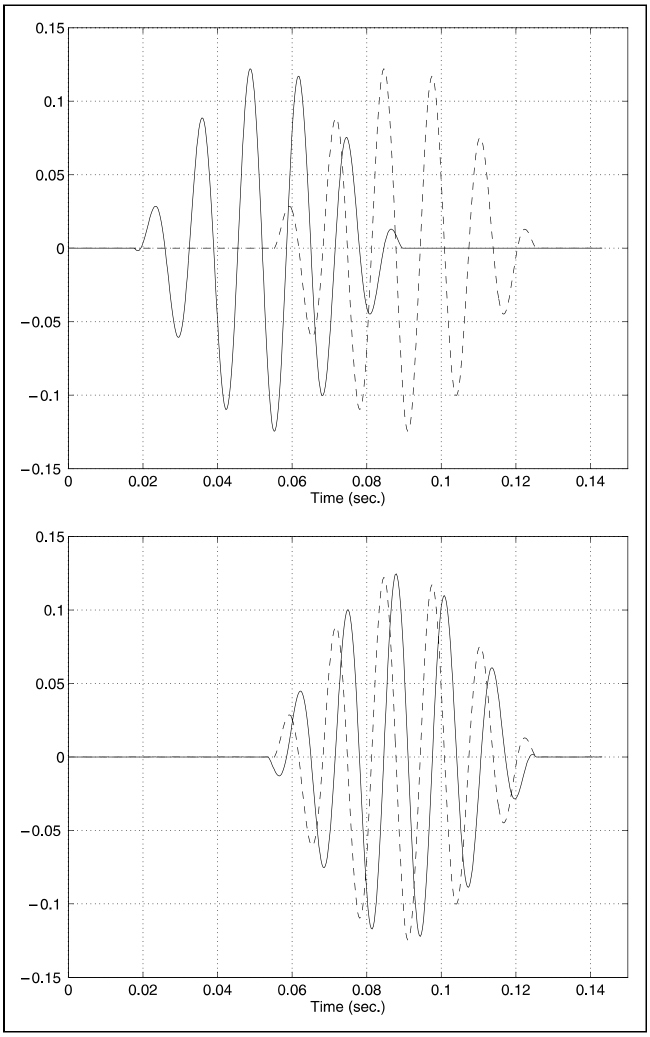

Figure 1: On the Top: Adjacent (orthogonal) cosine waveforms with smooth window $\mathbf{C}_{(2,1)}(t)$ and $\mathbf{C}_{(2,2)}(t)$. On the bottom: biorthogonal smooth local sine and cosine function $\mathbf{S}_{(2,2)}(t)$ and $\mathbf{C}_{(2,2)}(t)$.

the choice of the basis in order to approximate signals with sparse data sets. Sparse data often occur in classification problems and in the modelling of the control system, an exhaustive approach was given in [27].

Many applications in signal detection and image processing call for the use of functions that are local in time (or space) and frequency. The reason is that most signals have both temporal and spectral correlation. Therefore, the use of functions that are local in time and frequency is a good idea in order to point out the characteristics of the signal. The biorthogonal smooth local trigonometric bases have basically fast convergence and good approximation level [12]. In particular, the smoothness of the basis guarantees a continuous approximation, see [13]. In Fig. 1 biorthogonal smooth local sine and cosine functions are depicted. Normally when one is talking about the trigonometric wavelet functions, it is not immediately clear how to best divide the real axis into intervals.

One proposes an adaptive algorithm based on the minimization of Shannon's entropy function which splits the time interval depending on the analyzed signal [5]. One is talking about smooth localized trigonometric frames associated to covering by intervals of $\mathcal{R}$ as in our case or, more in general, of a manifold.

Definition 2. *Let a library of wavelet packets be the collection of functions of the form*

$$\psi_{(d,j,n)}(t) = \psi_j(2^d t - n) \tag{3}$$

where $(d, n) \in Z$ *and* $j \in N$.

It has already been remarked that one is talking about truncated indices, thus finite libraries of wavelet packets. Here, the *pyramidal* packet is represented by the indices (d, j, n), d is the level of the tree (scaling parameter), j is the frequency cell (oscillation parameter) and n the time cell (localization parameter).

The function $\psi_{(d,j,n)}(t) = \psi_j(2^d t - n)$ is roughly centered at $2^{-d}n$, has support of size $\approx 2^{-d}$ and oscillates $\approx j$.

3 Analytic Problem Formulation

Classification problems can be made to look like nonparametric regression if the outputs of the estimated function (network) are interpreted as being proportional to the probability that the input belongs to the corresponding class.

As already mentioned above, the practical problem which one wants to pose, is to reliably identify as soon as possible a particular transient which normally arises in the rail line system, *transformer inrush current*. Thus, one is looking for a *precise and fast algorithm* which can be implemented to run in real time. In classification problems, for training wavelets, sometimes a criterion based on the entropy is used, see [2; 5; 15]. This gives the *maximum-likelihood* estimate of the probability [2].

The entropy function is defined as

$$\mathcal{V} = -\sum_{\phi \in \mathcal{C}} \frac{\hat{\mathcal{P}}(\phi)}{\mathcal{P}} \ln\left(\frac{\hat{\mathcal{P}}(\phi)}{\mathcal{P}}\right), \tag{4}$$

where the $\hat{\mathcal{P}}$ and $\mathcal{P}$ are the estimated and the true probability respectively for a given signal to belong to the class $\mathcal{C}$. More, the $\mathcal{V}$ entropy function is the measure of the information needed to locate a system in a certain state, in other words $\mathcal{V}$ is the measure of the ignorance about the system. In our problem this function plays a special role. In fact, one wants to remember that the $\mathcal{V}$ entropy function of the expansion can be seen as a distance which measures the efficiency of a particular basis for expanding a given function. Roughly speaking, one basis is more efficient than another if its coefficients decrease to zero more rapidly, this means in fact that the probability to find the signal *at the time 't'* in this subspace is higher than another. In our problem, this function seems to play a particular role because, as already explained one wants to classify as soon as possible and with the best precision the signals which arise from the rail network: one needs the subspaces which correspond to the best compression data!

3.1 Selecting Best Wavelet Regressors

Selecting the 'best' regressors from a finite set of regressor candidates is a typical problem in regression analysis [10]. In our case, the sets of regressor candidates are two truncated smooth trigonometric wavelet library frames $\{\mathbf{C}_{(d,j,n)}(t), \mathbf{S}_{(d,j,n)}(t)\}$. The problem is then to select a number of elements from $\{\mathbf{C}_{(d,j,n)}(t), \mathbf{S}_{(d,j,n)}(t)\}$ based only on the output training data $\mathcal{O} = \{\mathbf{Y}_1(t), \mathbf{Y}_2(t), \ldots, \mathbf{Y}_N(t)\}$ in order to build the regression:

$$f(t) = \sum_{\psi_{(d,j,n)} \in \{\mathbf{C}_{(d,j,n)}, \mathbf{S}_{(d,j,n)}\}} \mathbf{u}_{(d,j,n)} \psi_{(d,j,n)}(t)\,. \tag{5}$$

More formally, our problem could be formulated as follows:

Regressor selection

Given a set of output observing data

$$\mathcal{O} = \{\mathbf{Y}_1(t), \mathbf{Y}_2(t), \ldots, \mathbf{Y}_N(t)\}$$

and the truncated wavelet frames:

$$\left\{\mathbf{C}_{(d,j,n)}(t), \mathbf{S}_{(d,j,n)}(t);\, (d,n) \in Z,\, j \in N,\, t \in \Re\right\}. \tag{6}$$

Then, find the elements belonging to the frame (6) *in the weights* $\mathbf{u}_{(d,j,n)}$ *according to the index*

$$\mathbf{J}\left(\psi_{(d,j,n)}\right) = \frac{1}{N}\sum_{k=1}^{N}\left(\mathbf{Y}_k(t) - \sum_{\psi_{(d,j,n)} \in \{\mathbf{C}_{(d,j,n)}, \mathbf{S}_{(d,j,n)}\}} \mathbf{u}_{(d,j,n)}(k)\psi_{(d,j,n)_k}(t)\right)^2 \tag{7}$$

which minimize $\mathcal{V}$ *as defined in* (4)*. The true probability is defined as the norm of the signal on the original subspace* $\mathcal{P} = \|\mathbf{Y}_k(t)\|^2$ *and the estimated probability as the norm of the signal on the approximating subspace* $\hat{\mathcal{P}} = \|\mathbf{Y}_k(t)\|^2$*. In other words:*

$$\hat{\mathcal{P}} = \left\| \sum_{\psi_{(d,j,n)} \in \{\mathbf{C}_{(d,j,n)}, \mathbf{S}_{(d,j,n)}\}} \mathbf{u}_{(d,j,n)}(k)\psi_{(d,j,n)_k}(t) \right\|^2 .$$

Remark 1. *Note that the procedure of minimization in* $\mathbf{u}_{(d,j,n)}$ *leads, at the first step, to the least squares solution with respect to* $\mathbf{u}_{(d,j,n)}$ *for every time-frequency cell and then, at the second step, to the minimization of the* $\mathcal{V}$ *function. Moreover, it is easy to remark that the maximum for the estimated probability* $\hat{\mathcal{P}}$ *is the minimum for the index* (7) *previously defined because the elements* $\mathbf{u}_{(d,j,n)}$ *are calculated according to the least squares method. The problem consists of choosing the expansion whose* information cost *and* error cost *are smallest.*

4 Several Basic Issues

The first question is which function can be used as activation function. The collected experience on this sense doesn't help too much. All of the model structures are capable of approximating any 'reasonable function' [7]. Thus, the question is to pick one that 'suits the application', in the sense that only few terms will be needed. A suitable criterion known in the literature is to select the basis which, once fixed a threshold level, has the minimum number of elements in the selected frame. Now, having chosen the best family, how to choose the size of the frame subset? Finally, how to select the terms of the subset?

4.1 Choosing the Best Regressor Family and the Adaptive Library

The case presented in this paper has quasi-harmonic signals that change amplitude and phase over time. This aspect suggests the wavelet like activation function. In Fig. 2 is depicted the windowing Fourier transform of a typical measured real signal. The data are very well concentrated around several frequencies, in this case integer multiples of the fundamental (50 Hz). The picture in Fig. 2 seems

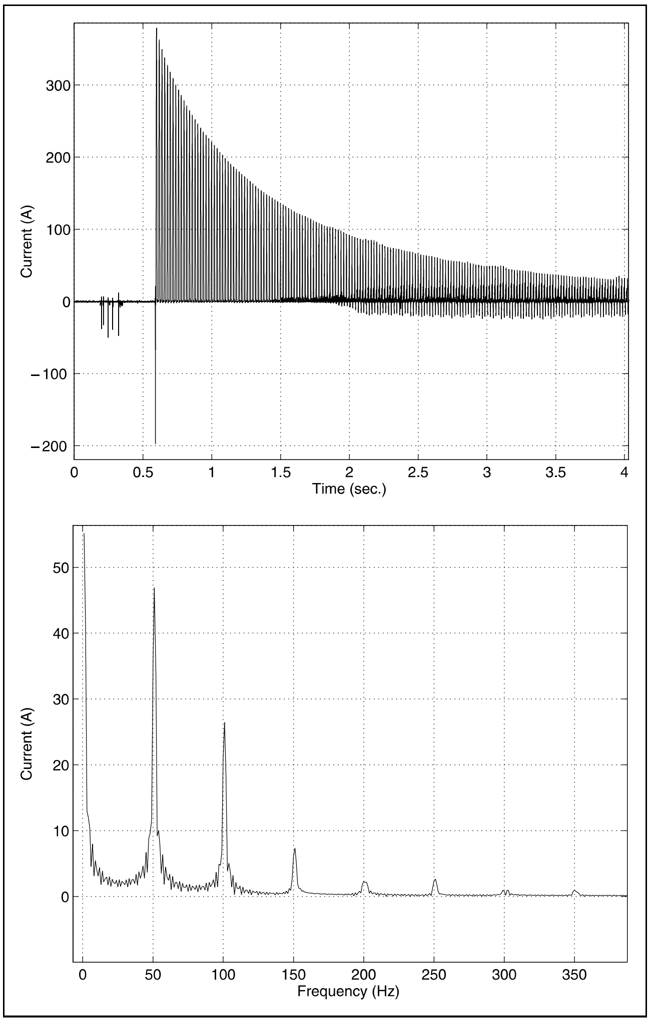

Figure 2: Real Signal. On the Top: Inrush current: time domain. On the Bottom: Inrush current: windowed spectral analysis.

to suggest a function with a frequency window and time support. Moreover, once selected the family regressor, for instance the truncated sine/cosine wavelets, the (d, j, n) parameterized family:

$$\mathcal{R}_c = \left\{ \psi^c_{(d,j,n)}(t);\ (d,n) \in Z,\ j \in N,\ t \in \mathfrak{R} \right\}$$

and

$$\mathcal{R}_s = \left\{ \psi^s_{(d,j,n)}(t);\ (d,n) \in Z,\ j \in N,\ t \in \mathfrak{R} \right\}$$

should contain a finite number of wavelets, as few as possible, so that the regressor selection procedure can efficiently be applied.

Given an approximating wavelet library not all the wavelet functions are useful, normally only a small number of the coefficients are important, the other ones can be neglected.

In order to explain the construction of wavelet network let us start with a regular wavelet lattice. Many wavelets in the regular lattice do not contain any data point in their support because of the sparseness of the data. The training data points don't provide any information for determining the coefficients of these empty wavelets, which means that they are superfluous for the regression estimation and could be eliminated. In general, one can select the candidate library as follows:

$$\left\{ \mathcal{R}_c,\ \mathcal{R}_s \right\} = \left\{ \psi^c_{(d,j,n)}(t), \psi^s_{(d,j,n)}(t) : (d,\ j,n) \in \mathcal{I}_1 \cup \mathcal{I}_2 \cup \ldots \cup \mathcal{I}_K \right\} \tag{8}$$

with $K = 1, 2, \ldots, L$.

and

$$\mathcal{I}_k = \left\{ (d,\ j,n) : \|\psi^c_{(d,j,n)}\|_p > \epsilon,\ \|\psi^s_{(d,j,n)}\|_p > \epsilon \right\}, \tag{9}$$

where ϵ is a chosen small positive number. In this way the 'empty' wavelets are eliminated from the wavelet frame. In other words, one starts from a regular tree packet (library) and only those which their support hit our training data are selected. This method is called *wavelet shrinkage* by some authors [3; 16].

One will show that with very few bases of the local trigonometric functions a good function detector can be obtained. If one sees the Fig. 2 it is easy to remark that our system could be described just around very few frequency with *suitable* dilation packets.

Another important motivation that can generalize this criterion will be more clear when one will describe the algorithm, in fact one computational efficient way to minimize the function (4) is to use biorthogonal frames.

4.2 Figures of Merit and Number of Wavelets

In order to evaluate the performance of nonparametric estimators, or to compare different nonparametric estimators or just different algorithms, it is necessary to introduce some figures of merit. As an evaluation of the estimator $f(t)$, one will adopt, the commonly used mean of square errors (**MSE**):

$$\mathbf{MSE} = \frac{1}{N} \sum_{k=1}^{N} (f(t) - y_k)^2 . \tag{10}$$

The **MSE** is a preliminary, even though necessary, index which tells about the error but it is insensitive with respect to the complexity of the estimator. In order to balance between the estimator errors and the complexity of the estimator, one has to consider other figures of merit which indicate the number of the wavelets of the network. The number of wavelets which one needs to be selected (i. e. the size of the wavelet network) is not an easy task. This normally corresponds to the usual model order determination, the problem is related to model structure or model order reduction. Even for linear systems, choosing a suitable model order is not a trivial problem. In order to recognize the signals one needs good *pattern recognition*, good features lead to good recognition. In this sense one has to be careful: it is known, in fact, that the result of a *bruteforce* procedure would give a function which perfectly fits the noisy data, which is known as 'overfitting' in the neural network literature. In other words this means that a suitable subspace exists and characterizes the physical signal. In literature several heuristic solutions can be found. In practice it is necessary to test two or three criteria. In general, these consist of introducing a penalty which is proportional to the model order. One has numerically tested two criteria for determining the number of wavelets in the wavelet network. This corresponds to determine the level of the tree in the pyramidal frame wavelet structure. One of these criteria is the Akaike's final prediction error criterion (**FPE**) [18]. This criterion is written as follows:

$$\mathbf{J}(f(t)) = \frac{1 + \frac{n_p}{N}}{1 - \frac{n_p}{N}} \frac{1}{N} \sum_{k=1}^{N} (f(t) - y_k(t))^2 , \tag{11}$$

where $f(t)$ is the wavelet network (the regression), n_p is the number of the parameters in the network estimator, $y_k(t)$ are the training data and N is the sample length of the training data. According to the notation introduced in (3) each wavelet has 1 dilation, n translation and 1 weight. In addition, the network has $n+1$ direct linear connections which leads to the following expression:

$$n_p = M(n+2) + n + 1 ,$$

for further details in [18].

Another criterion which is often used is the *Generalized Cross Validation* (**GCV**) which consists of estimating the expected performance with fresh data.

$$\mathbf{GCV} = \sum_{k=1}^{N} (f(t) - Y_k(t))^2 + \frac{2s\sigma^2}{N} , \tag{12}$$

where $f(t)$ is the wavelet network (the regression), s is the number of the wavelets in the network and N is the sample length of the training data. These indices shows how many wavelets should be included in the network, in other words how many levels the pyramid packet frame structure should have.

5 A Dual Lattice for Training the Wavelet Networks

The neural networks which use the family of wavelet functions as activation functions can be considered as a particular case of radial basis function networks [23]. The main objective of network training is to build a statistical model of the data generation process. Recent approaches [14] have developed efficient training techniques with a statistical point of view.

As mentioned above, the algorithm can be applied to every signal even though its effectiveness is strongly emphasized on narrow band signals or on signals which can be represented with a limited number of harmonics. Normally one layer is almost enough in the totality of the practical cases. In fact, with many basis functions, one hidden layer network is sufficient for modelling most practically reasonable systems, see [1; 6] or [15]. Practically reasonable systems are thus sufficiently approximated by only one-hidden layer even though in [24] the importance of the presence of the second hidden layer for the nonlinear systems is very well shown.

5.1 The Proposed Algorithm

One will give a preliminary description of the proposed algorithm, for further detail see subsection Sect. 5.2. Once the depth of the tree has been fixed by using the least squares method, the basic idea is to calculate the components on this frame and select a particular basis or frame looking for the minimum entropy level for each training signal. The idea is to decompose the residual signal of the previous approximation alternatively on the cosine and sine frame wavelet packets in order to minimize the entropy level for every residual approximation at every step. In order to consider and to use the nonorthogonality of the frames which generates an interaction between the elements of the bases[1] the algorithm considers at every step all the elements of the bases previously selected, without any elimination, see [27].

One can stop the algorithm with a threshold criterion at the stage i for the $\mathcal{L}^2$ norm of the differential error or, like in [14], for the entropy function.

The biorthogonal frames guarantee that at every step the residual signal approximation is performed in an orthogonalized subspace without constraint on the level of the packet tree which can be used for minimizing the Shannon entropy function.

[1] In a frame the decomposition is not unique.

Moreover, the structural biorthogonality of the bases guarantees robustness and efficient numerical calculations avoiding ill conditioned problems.

To be more precise:

$$\mathcal{J}\left(\mathbf{c}_{(d,j,n)}, \mathbf{s}_{(d,j,n)}\right) = \frac{1}{N}\sum_{k=1}^{N}\Bigg(\mathbf{Y}_k(t) - \sum_{(d,j,n)\in\mathcal{R}_c} \mathbf{c}_{(d,j,n)}\psi^c_{(d,j,n)}(t) - \sum_{(d,j,n)\in\mathcal{R}_s} \mathbf{s}_{(d,j,n)}\psi^s_{(d,j,n)}(t)\Bigg)^2 \qquad (13)$$

find the coefficients $\mathbf{c}_{(d,j,n)}$ and $\mathbf{s}_{(d,j,n)}$ belonging to the truncated cosine and sine frames represented as follows:

$$\mathbf{c}_{(d,j,n)} = \frac{\sum_{(d,j,n)\in\mathcal{R}_c}\left\langle \gamma_{c_{(i-1)}}(k), \psi^c_{(d,j,n)}(k)\right\rangle}{\left(\sum_{(d,j,n)\in\mathcal{R}_c}\left(\psi^c_{(d,j,n)}(k)\right)^2\right)}$$

and

$$\mathbf{s}_{(d,j,n)} = \frac{\sum_{(d,j,n)\in\mathcal{R}_s}\left\langle \gamma_{s_{(i-1)}}(k), \psi^s_{(d,j,n)}(k)\right\rangle}{\left(\sum_{(d,j,n)\in\mathcal{R}_s}\left(\psi^s_{(d,j,n)}(k)\right)^2\right)}$$

which minimize the $\mathcal{V}$ as defined in (4), more details in Sect. 5.2.

Remark 2. *One remarks that the coefficients* $\mathbf{c}_{(d,j,n)}$ *and* $\mathbf{s}_{(d,j,n)}$ *are constant on each own interval, which is connected with Heisenberg's indetermination problem. It is known that given a library of bases the entropy minimization criterion looks at the most equilibrated basis* [5], *now because of the biorthogonality of the sine and cosine frames the optimum can be found separately.*

The problem so presented suggests an interesting algorithm. In fact, because of the biorthogonality of our frames the minimum of the index, as defined in (13), can be found separately for each frame (Sine/Cosine). This could be performed in a dual approximation in two steps: the first step consists of the approximation on the cosine subspaces looking for *the most equilibrated basis* (minimum for the Shannon function) on this frame; the second step with the residual of this approximation again on the sine subspaces and so on recursively. Likewise, in order to choose the best basis for Shannon's entropy function one can use the fast and efficient algorithm already available in [5]. To conclude one remarks again that one can find the expansion whose *information cost* and *error cost* is smallest. The described technique processes the signals and in the mean time extracts the compressed data with the sine and cosine wavelet packet coefficients. The loop is stopped with a classical threshold technique performed on the difference of the norms between the signal and its approximation, when one will speak about the pattern recognition one will comeback to the argument.

One will give now a mathematical detailed description of the algorithm.

5.2 Mathematical Details

The above mentioned algorithm can be mathematically represented as follows.

Let

$$\mathcal{R}_c = \left\{ \psi^c_{(d,j,n)}(t);\ (d,n) \in Z,\ j \in N, t \in \Re \right\}$$

and

$$\mathcal{R}_s = \left\{ \psi^s_{(d,j,n)}(t);\ (d,n) \in Z,\ j \in N, t \in \Re \right\}$$

be the truncated cosine and sine packet frames respectively as defined in (8).

0. Define the initial residual $\gamma_{c(0)}(k) = \mathbf{Y}_k$, $k = 1, 2, \ldots, N$. Where the $\mathbf{Y}_k$ are the output observations in $\mathcal{O}$ as defined in Sect. 3. Fixed a *stage index* M such that, M minimizes the residual $\mathcal{L}^2$ norm as above mentioned, let $\mathbf{f}_0(t) = 0$.

Begin-loop

1. For $i = 1, 2, \ldots, M$.

Calculate the weights $\mathbf{c}_{(d,j,n)}(k)$ on all cosine wavelet packet tree according to the index:

$$\mathbf{J}_c\left(\mathbf{c}_{(d,j,n)}(k)\right) =$$

$$\frac{1}{N}\sum_{k=1}^{N}\left(\gamma_{c_{(i-1)}}(k) - \sum_{(d,j,n)\in\mathcal{R}_c} \mathbf{c}_{(d,j,n)}(k)\psi^c_{(d,j,n)}(k)\right)^2$$

this yields:

$$\mathbf{c}_{(d,j,n)} = \frac{\sum_{(d,j,n)\in\mathcal{R}_c}\left\langle \gamma_{c_{(i-1)}}(k),\ \psi^c_{(d,j,n)}(k)\right\rangle}{\left(\sum_{(d,j,n)\in\mathcal{R}_c}\left(\psi^c_{(d,j,n)}(k)\right)^2\right)},$$

where $\gamma_{(i-1)}(k)$ $(k = 1, 2, \ldots, N)$ are the residuals of the stage $(i-1)$.

2. Let

$$\mathcal{V} = -\sum_{k=1}^{N}\left(\frac{\hat{\mathcal{P}}\left(\gamma_{c_{(i-1)}}(k)\right)}{\mathcal{P}\left(\gamma_{c_{(i-1)}}(k)\right)}\right)\ln\left(\frac{\hat{\mathcal{P}}\left(\gamma_{c_{(i-1)}}(k)\right)}{\mathcal{P}\left(\gamma_{c_{(i-1)}}(k)\right)}\right),$$

where the true probability is

$$\mathcal{P}\left(\gamma_{c(i-1)}(k)\right) = \|\gamma_{c(i-1)}(k)\|^2$$

and the estimated probability is:

$$\hat{\mathcal{P}}\left(\gamma_{c(i-1)}(k)\right) = \left\| \sum_{(d,j,n)\in\mathcal{R}_c} \mathbf{c}_{(d,j,n)}(k)\psi^c_{(d,j,n)}(k)\right\|^2.$$

$$arg(min_{\mathcal{R}_c}(\mathcal{V})) = \left\{ l_{c_{(d,j,n)}} \right\} \text{ with } (d,\ j,\ n) \in \{\mathcal{R}_c\}\ .$$

(This step selects the adaptive dilation on the cosine frames).

3. Update $\mathbf{f}(t)$ and γ:

$$\mathbf{f}_i(t) = \mathbf{f}_{(i-1)}(t) + \sum_{l_c\in\mathcal{R}_c} \mathbf{c}_{l_{c_{(d,j,n)}}}\psi^c_{l_{c_{(d,j,n)}}}(t)$$

$$\gamma_{c_i}(k) = \gamma_{c_{(i-1)}}(k) - \sum_{l_{c_{(d,j,n)}}\in\mathcal{R}_c} \mathbf{c}_{l_{c_{(d,j,n)}}}(k)\psi^c_{l_{(d,j,n)}}(k)\ ;$$

$$k = 1, 2, \ldots, N\ .$$

4. Calculate the weights $\mathbf{s}_{(d,j,n)}(k)$ on all sine wavelet packet tree according to the index:

$$\mathbf{J}_s\left(\mathbf{s}_{(d,j,n)}(k)\right) =$$

$$\frac{1}{N}\sum_{k=1}^{N}\left(\gamma_{s_{(i-1)}}(k) - \sum_{(d,j,n)\in\mathcal{R}_s} \mathbf{s}_{(d,j,n)}(k)\psi^s_{(d,j,n)}(k)\right)^2$$

this yields:

$$\mathbf{s}_{(d,j,n)} = \frac{\sum_{(d,j,n)\in\mathcal{R}_s}\left\langle \gamma_{s_{(i-1)}}(k),\ \psi^s_{(d,j,n)}(t)\right\rangle}{\left(\sum_{(d,j,n)\in\mathcal{R}_s}\left(\psi^s_{(d,j,n)}(t)\right)^2\right)},$$

where $\gamma_{s_{(i-1)}}(k) = \gamma_{c_i}(k) \quad (k = 1, 2, \ldots, N)$.

5.

$$\mathcal{V} = -\sum_{k=1}^{N}\left(\frac{\hat{\mathcal{P}}\left(\gamma_{s_{(i-1)}}(k)\right)}{\mathcal{P}\left(\gamma_{s_{(i-1)}}(k)\right)}\right)\ln\left(\frac{\hat{\mathcal{P}}\left(\gamma_{s_{(i-1)}}(k)\right)}{\mathcal{P}\left(\gamma_{s_{(i-1)}}(k)\right)}\right),$$

where the true probability is

$$\mathcal{P}\left(\gamma_{s_{(i-1)}}(k)\right) = \|\gamma_{s_{(i-1)}}(k)\|^2$$

and the estimated probability is:

$$\hat{\mathcal{P}}\left(\gamma_{s_{(i-1)}}(k)\right) = \left\| \sum_{(d,j,n)\in\mathcal{R}_s} \mathbf{s}_{(d,j,n)}(k)\psi^s_{(d,j,n)}(k)\right\|^2.$$

$$arg(min_{\mathcal{R}_s}(\mathcal{V})) = \left\{ l_{s_{(d,j,n)}} \right\} \text{ with } (d,\ j,\ n) \in \{\mathcal{R}_s\}\ .$$

(This step selects the adaptive dilation on the sine frames).

6. Update $\mathbf{f}(t)$ and γ:

$$\mathbf{f}_i(t) = \mathbf{f}_i(t) + \sum_{l_{s_{(d,j,n)}}\in\mathcal{R}_s} \mathbf{s}_{l_{s_{(d,j,n)}}}\psi^s_{l_{s_{(d,j,n)}}}(t)$$

$$\gamma_{s_i}(k) = \gamma_{s_{(i-1)}}(k) - \sum_{l_{s_{(d,j,n)}}\in\mathcal{R}_s} \mathbf{s}_{l_{s_{(d,j,n)}}}(k)\psi^s_{l_{s_{(d,j,n)}}}(k)$$

$$k = 1, 2, \ldots, N\ .$$

$$\gamma_{c_i}(k) = \gamma_{s_{(i)}}(k)\ (k = 1, 2, \ldots, N)\ .$$

End Loop.

This algorithm seems to have very good characteristics, in fact, because of the biorthogonality the approximating subspaces are structurally orthogonal to each other and the procedure do not need any orthogonalization. Moreover, the unitary bases allows a procedure without any matrix inversion.

6 Applications

In this section one will show the effectiveness of the proposed algorithm, in particular our attention is directed to a classification of inrush current in rail vehicle systems. The simulation presented was built with a software platform by using the toolbox in [9]. One will start with several preliminary considerations.

6.1 Choosing the Number of Wavelets and Pattern Recognition

Considering the indices (11) and (12) one can choose the number of wavelets for the frame packet. Three possible libraries are been calculated, depth of the tree equal 2, 3 and 4. The indices (11) and (12) suggest to choose the level with depth equal 2. The data are split in two parts, 20 old data to build the model and 20 fresh data to validate it.

A statistical pattern recognition every 40 ms (the time resolution for the deepest basis) is built. The pattern recognition is based on the compressed data, the unique prototype constructed from the training data is depicted in Fig. 3. The network is tested with two different matching functions: the first one is the difference between the euclidean norm of the projected and compressed fresh signal on our approximating subspace and the norm of the compressed prototype signal; the second one is the difference between the entropy level of the compressed prototype signal and the entropy level of the projected and compressed fresh signal on our approximating subspace, like in [14]. Anyhow the problem is a classification problem with two classes: *Inrush / No Inrush*.

The results obtained with the two criteria are qualitatively the same, for sake of brevity one will report and discuss the first one. In other words, with the first criterion one is looking for the *statistical distance* from the subspace which minimizes the *Shannon entropy function*. The euclidean norm of this distance, cut by a statistical threshold level, allows to classify the signals. Regarding the data compression several words should be spent in order to reach a better understanding of the meaning. Once the prototype function is built from the training data, if is compressed with a cosine/sine basis in a vector with four coefficients for every time cell, corresponding to the [0; 50; 100; 150] Hz. The compression procedure generates vectors of a length of not more than 16 components, 4 components for every time cell. The inner product between the prototype vector and the fresh data vector needs mostly only one time-frequency cell to recognize the inrush (40 ms).

6.2 Results

In the preliminary simulation 20 signals are considered as training signals and another 20 signals as fresh testing signals. One can see how the effect of the algorithm with sine/cosine frames is particularly *selective*. It is known, in fact, from all the neural network literature, that a good algorithm should create well separated class conditional probability distributions. If the distributions are well separated, discrimination is simple. Otherwise a very high

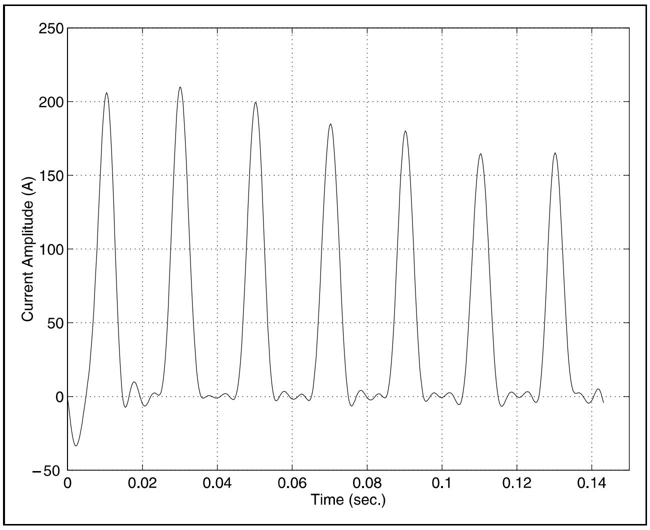

Figure 3: Prototype.

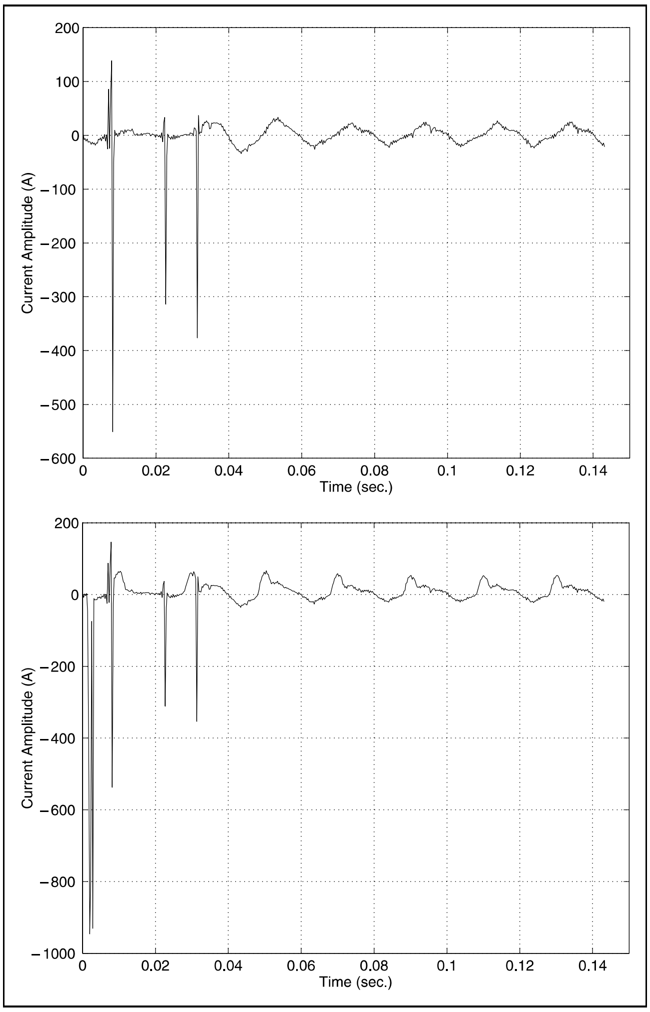

Figure 4: On the Top: Typical disturbance in the neutral section. On the Bottom: Inrush current and typical disturbance in superposition (the worst case).

percentage level of wrong identification occurs, in particular the network cannot distinguish the noise from the signal. It is also known that the distribution of the training data set plays an important role. In our case, selecting a wider amplitude distribution of the training signals, one achieves 100% correct results. In Fig. 4 one of the typical real disturbances is depicted. The worst case happens when the superposition of real noise meets an inrush current with low amplitude as depicted in Fig. 4, this was recognized in 120 ms instead of 40 ms. Because of the similarity between network noise and the inrush current it is not easy task to recognize the inrush from the noise. In particular, on the top of the Fig. 4 a typical network noise is depicted and on the bottom the inrush current with the network noise in superposition.

7 Conclusion

A technique of nonparametric regression and classification with wavelet networks is proposed. An algorithm is built for neural network training and filtering based on recursive iterations over dual smooth trigonometric frames in order to detach, compress and classify the signals. The developed algorithm combines techniques in regression analysis and backpropagation procedures. It consists of recursive dual iterations with biorthogonal smooth local sine and cosine wavelet packets in order to fit the training data signal set with the "best Shannon basis" and to choose the expansion whose *information cost* is smallest. In order to consider and to use the nonorthogonality of the frames, which generates an interaction between the elements of the bases, the algorithm considers at every step again all the elements of the bases previously selected. Moreover, the structural biorthogonality of the frames guarantees robustness and efficient numerical calculations. The algorithm is totally general even though it is particularly effective for narrow band signals or for harmonics detection. Speed and precision are achieved with a limited number of the wavelets.

Acknowledgements

We want to thank European Commission for the economical support and ABB Corporate Research (Heidelberg) for the technical support.

References

[1] A. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory*, **IT-39**, pages 930–945, 1993.

[2] E. Baum, F. Wilczek. Supervised learning of probability distributions by neural networks. *Neural Information Processing Systems. American Institute of Physics*, pages 52–61, 1988.

[3] A. Benveniste, A. Juditscky, B. Delyon, Q. Zhang, P.G. Glorenee. Wavelets in identification. In *Proc. SYSID'94 10th IFAC Symp. Syst. Identification*, Copenhagen, 1994.

[4] M.A. Cody. The wavelet packet transform, extending the wavelet transform. *Dr. Dobb's Journal*, 19:44–46, 1994.

[5] R.R. Coifman, M.V. Wickerhauser. Entropy based algorithm for best basis selection. *IEEE Trans. Inform. Theory*, 32:712–718, 1992.

[6] G. Cybenko. Approximation by superposition of a sigmoidal function. *Math. Control, Signal and Syst.*, (2):303–314, 1989.

[7] I. Daubechies. The wavelets transform, time frequency. *IEEE Trans. Inf. Theory*, **IT-36**, pages 961–1005, 1992.

[8] I. Daubechies. *Ten Lectures on Wavelets*. Publisher Society for Industrial and Applied Mathematics, Philadelphia (Pennsylvania), 1995.

[9] D.L. Donoho. http://playfair.stanford.edu/ wavelab. 1996.

[10] N. Draper, H. Smith. *Applied Regressor Analysis*. 2nd ed. Wiley New York, 1981.

[11] R.J. Duffin, A.C. Schaeffer. A class of nonharmonic fourier series. *Trans. Amer. Math. Soc.*, (72):341–366, 1952.

[12] X. Fang and E. Séré. Adaptive multiple folding local trigonometric transforms and wavelet packets. *Appl. Comput. Harmon. Anal.* 1(2), 1994.

[13] B. Jawerth, W. Sweldens. Biorthogonal smooth local trigonometric bases. *J. Fourier Anal. Appl.*, 2(2):109–133, 1995.

[14] M. Joost, W. Schiffmann. Speeding up backpropagation algorithms by using cross-entropy combined with pattern normalization. *IJUFKS-Journal*, 6(2):117–126, 1999.

[15] A. Juditscky, H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg, Q. Zhang. Nonlinear black-box models in system identification: mathematical foundetion. *Automatica*, (31):1725–1750, 1995.

[16] A. Juditscky, Q. Zhang, B. Delyon, P.Y. Glorennec, A. Benveniste. Wavelets in identification. *Thecnical Report, IRISA*, 1994.

[17] K. Gröchening. Acceleration on the frame algorithm. *IEEE Trans. on Signal Processing*, 41(12), 1993.

[18] L. Ljung. *System identification: Theory for the User*. Information and System Science Series, Prentice Hall., 1987.

[19] S. Mallat, Z. Zhang. Matching pursuit with time-frequency dictionaries. *Technical Report 619, New York University, Computer Science Department*, 1993.

[20] P. Mercorelli, M. Rode, P. Terwiesch. A wavelet packet algorithm for online detection of pantograph vibrations. In *Proc. 9th IFAC Int. Symp. on Control in System Transportation 2000*, Braunschweig, 2000.

[21] P. Mercorelli, P. Terwiesch. A black box identification in harmonic domain. *VDE European Transactions on Electrical Power, to appear as regular paper*, 2002.

[22] B. Perunicić, M. Mallini, Z. Wang, Y. Liu. Power quality disturbance detection and classification using wavelets and artificial neural networks. In *Proc. 8th IEEE International Conference on Harmonics and Quality of Power,* Athens 1998.

[23] M. Pottmann, D.E. Seborg. Identification of nonlinear process using reciprocal multiquadratic functions. *J. Proc. Control*, 2, 1992.

[24] E. Sontag. *Neural network for control*. Publisher H. Trentelmann and J. Willems, Boston, 1993.

[25] P. Terwiesch, S. Menth, S. Schmidt. Analysis of transients in electrical railway networks using wavelets. *IEEE Trans. on Industrial Electronics*, 45(6):955–959, 1998.

[26] P. Terwiesch, P. Mercorelli. A local feature extraction using biorthogonal bases for classification of embedded classes of signals. In *Proc. 14th Int. Symp. on Mathematical theory of networks and systems (MTNS 2000)*, Perpignan (France), 2000.

[27] Q. Zhang. Using wavelet network in nonparametric estimation. *IEEE Transaction on Neural Networks*, 8(2):227–236, 1997.

Manuscript received: 20th August 2001.

Dr.-Ing. Paolo Mercorelli is with IAI in Wernigerode (Germany) as industrial researcher since March 2002. His actual interests are: geometric control, robotic manipulation, wavelets, sliding mode control, flatness based control.

Address: Institut für Automatisierung und Informatik GmbH, Schlachthofstr. 4, D-38855 Wernigerode, Tel: +49-(0)3943-6259746, Fax: +49-(0)3943-6259747, E-Mail: mercorelli@iai-wr.de

Dr.-Ing. Peter Terwiesch is the head of ABB's Process Industries division in Switzerland and technology head for the global Process Industries division. He is serving on a number of boards, including those for VDI/VDE GMA and for at-Automatisierungstechnik.

Address: Automation Technologies – ABB Switzerland Ltd, Process Industries, CH-5401, Baden-Dättwil, Tel: +41 585 86 88 88, Fax: +41 585 86 73 11, E-Mail: Peter.Terwiesch@ch.abb.com

Prof. Domenico Prattichizzo is Associate Professor in Automatic Control at the Siena University. He is Co-Chair of the 2nd IEEE CSS and RAS Int. Workshop on Control Problems in Robotics and Automation Las Vegas, Dec. 2002. His current research activities include robotic manipulation, mobile robotics, vision, optimal control.

Address: Dip. di Ingegneria dell' Informazione – Università di Siena, via Roma, 56, 53100 Siena, Italia, Tel: +39-0577-234609, Fax: +39-0577-233602, E-Mail: prattichizzo@ing.unisi.it