THE USE OF WAVELET PACKETS FOR DETECTION AND IDENTIFICATION OF EVENTS IN UTERINE EMG

M. Chendeb*, M. Khalil** and J. Duchêne*

* University of Technology of Troyes, ISTIT, 12 rue Marie Curie, BP 2060, 10010 Troyes, France
** Lebanese University, Faculty of Engineering, Electronic Department, Tripoli, Lebanon

marwa.chendeb@utt.fr

Abstract: Uterine EMG is an efficient tool to follow the evolution of contractility during pregnancy. Its analysis requires efficient tools for detection and identification of events contained in the recordings. Four events have been identified in uterine EMG that can be used to assist in preterm birth diagnosis. The present work is based on the use of Wavelet Packet (WP) decomposition. In an unsupervised way, WP decomposition is used in association with the Kullback Leibler distance KLD, which provides a criterion related to detection capability. KLD is applied directly on the wavelet packet coefficients rather than on the reconstructed signals. When there is no event to be detected, the estimated KLD roughly follows an exponential distribution. As WP decomposition produces a redundant tree, a best basis selection is based on the suppression of WP without any specificity in terms of change detection. Results evidenced the efficiency of the method for simulated signals as well as for real uterine EMG recordings. Redundancy reduction suppressed half the number of WP selected firstly without any degradation of the overall detection performance. Any application where events to be detected are characterized by their frequency content is a good candidate for such a methodology.

Introduction

When recorded with abdominal surface electrodes, uterine EMG is an efficient tool to follow the evolution of contractility during pregnancy [1][2]. The non-invasive nature of this kind of recording makes it useful for pregnancy monitoring, particularly when associated with ambulatory instrumentation. Four events have been identified in uterine EMG that can be used to assist in preterm birth diagnosis: uterine contractions, foetus motions, Alvarez waves and long-duration low frequency band (LDBF) waves [2]. However these events are difficult to detect and isolate in recordings where the signal to noise ratio (SNR) is very poor.

The Wavelet Packet (WP) decomposition has been shown to be a very efficient tool for signal analysis and event detection. WP decomposition is a generalized version of the discrete wavelet transform that retains high and low-frequency sub-bands. The partitioning of the frequency axis may therefore take many forms to suit the needs of the application, and a powerful signal analysis on selected frequency sub-bands is obtained. The redundancy of information requires the choice of a best basis selection that represents at best the signal with respect to a given criterion. In the present work, the main goal is to choose the packets that reveal the different events that occur in the recordings. Therefore a criterion of packet selection has to be chosen according to this objective of event detection [3].

Wei and colleagues presented a study on active detection of delamination for multi-layer composites using a combination of modal analysis and WPT [4]. Peng and colleagues used WPT and an effective method for intrinsic mode function (IMF) selection in the rolling bearing fault detection [5]. The power of WPT is that a best basis can be chosen for a specific task, if it can be properly identified from the set of possible candidates. The choice of the basis depends on criteria applied by analysis goals, such as compression, filtering (smoothing) [6], feature extraction and classification [7][8], etc. Ravier and Amblard presented a detector of transient acoustic signals combining the local wavelet analysis and higher-order statistical properties of the signals [9]. Leman and Marque used WPT, and also proposed a more specific criterion to denoise the EHG signal [10]. Hitti and Lucas proposed a best basis selection method to detect abrupt changes in noisy multi-component signals [11]. They used energy criterion to allow separation of the different frequency components of the signal from a wavelet-packet library tree.

In this work, we propose a best basis selection to select a set of packets from the comprehensive wavelet packet tree. In an unsupervised way, WP decomposition is used in association with the Kullback Leibler distance KLD, which provides a criterion related to detection capability. KLD is applied directly on the WP coefficients rather than on the reconstructed signals. When there is no event to be detected, the estimated KLD roughly follows an exponential distribution. The Kolmogorov-Smirnov test is used to measure the distance between experimental and theoretical cumulative distributions to highlight the presence of ruptures, then to select the most relevant packets. The performance of the methods and examples using real datasets are also shown.

ISSN: 1727-1983

Materials and Methods

Wavelet Packet WP Decomposition: Wavelet packet decomposition is an extension of Discrete Wavelet Transform and can be obtained by a generalization of the fast pyramidal algorithm. Each detail coefficient vector is decomposed into two parts using the same approach as in approximation vector splitting. The complete binary tree is produced as shown in figure 1.

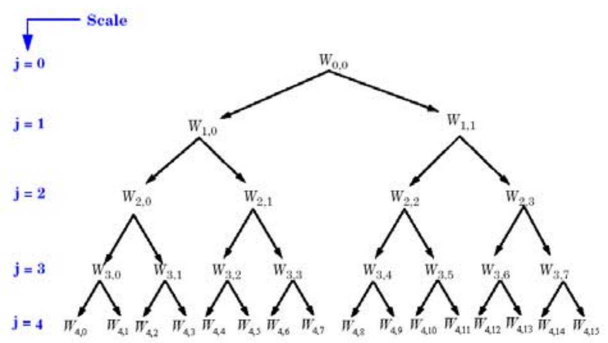

Figure 1: Wavelet Packet Decomposition Tree.

We start with *h(n)* and *g(n),* the two impulsive responses of low-pass and high-pass analysis filters, corresponding to the scaling function and the wavelet function, respectively. With each node is associated a subspace $\Omega_{j,n}$ generated by an orthonormal basis $\{\psi_{\mathrm{j,n}}(t-2^j n)\}_{n\in Z}$, j being interpreted as a scale parameter, n as a sequence parameter. The two bases at the children nodes (level (j $+1$)) are defined from the father nodes (level j) [12]:

$$\psi_{j+1,2n}(t)=\sum_{k=-\infty}^{+\infty} h(k)\psi_{j,n}(t-2^j k) \tag{1}$$

$$\psi_{j+1,2n+1}(t)=\sum_{k=-\infty}^{+\infty} g(k)\psi_{j,n}(t-2^j k) \tag{2}$$

Each packet is indexed by a subset of indices: *j, n* and *k* (time-localization index). The WP Coefficients at each node (*j, n*) are computed as:

$$\left(C_{j,n}(k)\right)=\left\langle f(t),\psi_{j,n}(t-2^j k)\right\rangle \tag{3}$$

$\left(C_{j,n}(k)\right)$ analyzes the fluctuations of the signal roughly around the position $2^j k$, at the scale 2^j and at various frequencies for the various admissible values of the parameter n [8].

Best basis selection: WP decomposition allows a well adapted analysis of a signal. In this step, the idea is to select a suitable orthogonal WP subset as a basis for further signal decomposition, according to the objectives of the expected signal analysis. For a *J* scale decomposition, the resulting binary tree yields $2^{J+1}-1$ packets offering a complete description of the space of the original signal. The set of subspaces in the binary tree is a redundant tree. To determine the best basis, a cost function must be chosen to represent the goal of the application. The commonly used criterion for choosing the most efficient or best basis for a given signal is the minimum entropy criterion [6].

Our goal is to define a criterion which permits the detection of the existence (or not) of events in the signal, then to allow classification of these events. A method to highlight the ruptures, hence to detect the presence of different events, is the use of the KLD. The KLD is directly applied on a temporal partition of the packet coefficients, and not on the reconstructed signals [3].

Cost function for detection purposes: As previously mentioned, our goal is to find a set of wavelet packets allowing detectability of different events in the signal. For detection problems, the cost function has to reveal the capacity of the WP to detect the presence of specific events (specific in the sense of the application characteristics). The Kullback Leibler distance has been already used as a discrimination measure, e.g., in classification problems [7], or image comparison [13][14]. In our approach we propose to use K-L distance as a basis for the definition of the cost function. It can be seen as a distance between two Probability Density Functions PDFs $f(X,\theta_i)$ and $f(X,\theta_j)$. The natural choice of D is the relative entropy (also known as cross entropy, Kullback Leibler distance, or I divergence) between two PDFs [7][15]:

$$D(f(X;\theta_i),(f(X,\theta_j))=\int f(x,\theta_i)\log\frac{f(x,\theta_i)}{f(x,\theta_j)}dx. \tag{4}$$

In fact this distance is accessible only by estimation. Providing two sequences $\mathbf{x_i}$ and $\mathbf{x_j}$, the estimation of the KLD between them implies the estimation of the distributions for each of them, so that KLD estimation necessitates the knowledge of the distribution of the WP Coefficients.

Experiments show that a good PDF approximation for the marginal density of wavelet packet coefficients at a particular sub-band transform may be achieved by adaptively varying the two parameters of the generalized Gaussian density (GGD) [16], which is defined as:

$$f(x;\alpha,\beta)=\frac{\beta}{2\alpha\Gamma(1/\beta)}e^{-(|x|/\alpha)^{\beta}} \tag{5}$$

where $\Gamma(.)$ is the gamma function: $\Gamma(z)=\int_0^{\infty} e^{-t}t^{z-1}dt$, $z\succ 0$.

 ISSN: 1727-1983

α is related to the width of the PDF peak (standard deviation), while β is inversely proportional to the decreasing rate of the peak. The GGD model is Gaussian when $\beta = 2$.

Using all available signals in a test set, it has been shown that uterine EMG amplitudes follow a Generalized Gaussian Distribution. Consequently, the packet coefficients follow the same distribution, as the WPT is a linear transformation (see figure 2). In the same way it has been shown that the "real" noise superimposed on these signals followed a normal distribution. Figure 2 shows a typical example of a histogram of the coefficients of a wavelet sub-band packet of uterine EMG and real noise, together with a plot of the fitted GGD using the Maximum Likelihood estimate of α and β [16]. As a result, any uterine EMG can be statistically described by a GGD, when the noise can be described by a simple Gaussian distribution ($\beta = 2$).

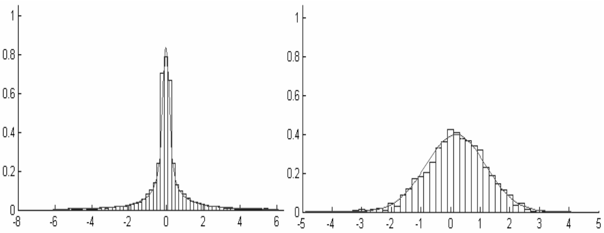

Figure 2: Wavelet sub-band coefficient histogram fitted with a generalized Gaussian density. Left hand drawing: normalized uterine EMG. Right hand drawing: normalized uterine recorded noise.

Distribution of the estimated KLD: If *N* is the length of the sequence $\mathbf{x} = \{x_1, x_2, ... x_N\}$, an estimation of the parameter is given as:

$$\hat{\alpha} = \sqrt{\left(\frac{2}{N}\sum_{i=1}^{N} x_i^2\right)} \tag{6}$$

The estimated KLD $\hat{K}$ between two PDFs from the Gaussian families with $\beta = 2$ and the estimated parameter α becomes:

$$\hat{K}_{ij} = \left[\log\left(\frac{\hat{\alpha}_i}{\hat{\alpha}_j}\right)^2 + \frac{1}{2}\left(\frac{\hat{\alpha}_j}{\hat{\alpha}_i}\right)^2 - \frac{1}{2}\right] \tag{7}$$

The KLD is not symmetric. To overcome this problem we use:

$$\hat{K} = \hat{K}_{ij} + \hat{K}_{ji} \tag{8}$$

$$\hat{K} = \frac{1}{2}\left[\left(\frac{\hat{\alpha}_i}{\hat{\alpha}_j}\right)^2 + \left(\frac{\hat{\alpha}_j}{\hat{\alpha}_i}\right)^2 - 2\right] \tag{9}$$

Let us also consider that only limited independent sequences $\mathbf{x_i} = \{x_{i1}, x_{i2}, ..., x_{iN}\}$ and $\mathbf{x_j} = \{x_{j1}, x_{j2}, ..., x_{jN}\}$ are available for K estimation.

$\hat{\alpha}_i$ and $\hat{\alpha}_j$ are computed using (6). $z = \left(\frac{\hat{\alpha}_i}{\hat{\alpha}_j}\right)^2$ and $\frac{1}{z} = \left(\frac{\hat{\alpha}_j}{\hat{\alpha}_i}\right)^2$ follow a $F_{N,N}$ Fisher distribution with $E(z) = \frac{N}{N-2}$ and $\mathrm{var}(z) = \frac{4N(N-1)}{(N-2)^2(N-4)}$.

The expectation of the estimated K-L distance $\hat{K}$ is [3]:

$$E(\hat{K}) = \frac{2}{N-2} \tag{10}$$

First order approximation of the $\hat{K}$ distribution: As we do not have an analytical expression of the distribution of $\hat{K}$, the idea is to approach this distribution with a known distribution having at least the same general shape and expectation. We constructed an empirical histogram of $\hat{K}$ by simulation (figure 3).

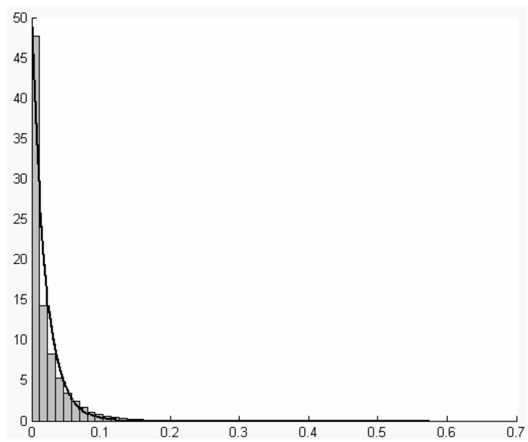

Figure 3: Histogram of $\hat{K}$ associated with the exponential distribution of same expectation. Simulation was made by generating 500 segments of Gaussian white noise (length N = 100).

Taking into account the shape of the obtained histogram, we chose as a first approximation of that histogram the exponential distribution, depending on

ISSN: 1727-1983 © 2005 IFMBE

only one parameter. Its probability density function is defined as:

$$f(x) = \lambda . e^{-\lambda . x} \quad \text{with} \quad E(x) = \frac{1}{\lambda} \tag{11}$$

An adjustment between the exponential distribution and the histogram is easily obtained by equalizing both expectations:

$$\frac{1}{\lambda} = \frac{2}{N-2} \tag{12}$$

Figure 3 shows an example of $\hat{K}$ histogram obtained by simulation from a Gaussian white noise $N(0,1)$, with a sequence length $N = 100$.

Wavelet Packet selection: The goal of this part is to retain only the wavelet packets that are able to detect changes in a specific class of signals (uterine EMG in our application). The main idea is that, if a wavelet packet contains at least one change, the distribution of $\hat{K}$ does not follow the exponential distribution with $\lambda = \frac{N-2}{2}$ any more.

As a criterion for WP selection, we used the Kolmogorov Smirnov statistics $D_{\max}$, distance between theoretical and sample cumulative distributions. At the moment, two approaches are to be imagined for WP selection: either defining a threshold on $D_{\max}$, or selecting $D_{\max}$ in descending order and limiting the number of selected WP to a predefined number.

Whatever the approach, a node in the WP tree will be put to “1” if the corresponding WP has been selected, the others being put to “0”.

The previous step identified all nodes where significant activities were detected. As the tree is highly redundant, the next steps have to select the nodes that will be finally kept for further signal analysis. The current implementation of the selection algorithm roughly follows the first proposed by Hitti and Lucas [11]. The steps of the algorithm selecting the best basis are the following:

a. The number 1 or 0 is associated with each packet according to the K-S result, with 1 meaning that there is at least one rupture (Fig.4a).

b. The value at each node father is compared with the sum of values of its sons. If the sum is larger than that of the father, the sum is then accorded to the father (Fig.4b).

c. Only the nodes at “1” having a father at “2” or higher than “2” are selected in order to reduce the redundancy (Fig.4c).

Hitti’s algorithm guarantees a complete basis representing the entire original signal (all the packets at 1 in Fig. 4c), i.e. the original signal can be reconstructed from the selected basis. Now, our goal is to select from this basis only the packets that are significant for event detection. According to this idea, we select only the packets at “1” in the first and third trees simultaneously. The final selected packets are those framed on Fig. 4c.

Change time detection: Any time-detection algorithm could work for the evaluation of the performance of the best basis selection algorithm. However, we had already developed a specific method well adapted to uterine EMG recordings, DCS (Dynamic Cumulative sum) [1]. Therefore we made use of this CUSUM like algorithm for detection purposes.

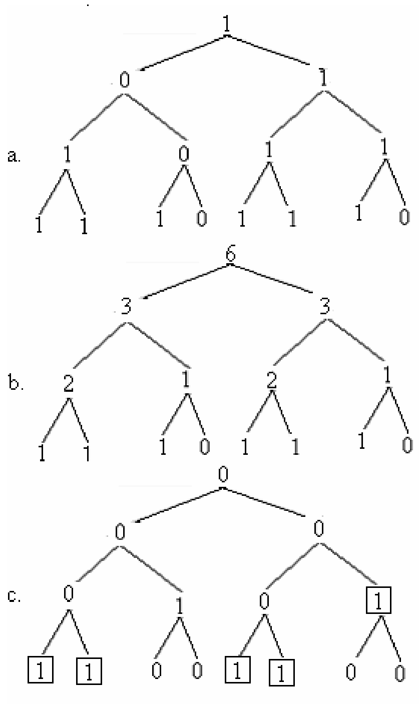

Figure 4: Steps for selection of the best basis.

In a few words, DCS works as follows: DCS is based on local cumulative sums of likelihood ratios computed between two locally estimated distributions around time t. The parameters of the distributions Θ_b (before) and Θ_a (after) are estimated using two windows of identical length before and after the current time t.

After parameter estimation, DCS is defined as a cumulative sum of likelihood ratios:

$$DCS(f_{\hat{\Theta}_a^t}, f_{\hat{\Theta}_b^t}) = \sum_{j=1}^{t} \log \frac{f_{\hat{\Theta}_a^j}(X_j)}{f_{\hat{\Theta}_a^j}(X_j)} \tag{13}$$

The detection function is defined as:

$$g(t) = \max_{1 \le j \le t} \left[DCS\left(f_{\hat{\Theta}_a^j}, f_{\hat{\Theta}_b^j} \right) \right] - DCS\left(f_{\hat{\Theta}_a^t}, f_{\hat{\Theta}_b^t} \right) \tag{14}$$

ISSN: 1727-1983 © 2005 IFMBE

Finally the stopping time is:

$$t_p = \inf\{n \geq 1 : g(t) > h\} \qquad (15)$$

where h is the threshold defined from a training set (ROC curve).

Uterine EMG signals: Uterine EMG were recorded on 32 women, hospitalized for risk pregnancy. The gestational age at recording ranged from 19 to 38 weeks. The sample frequency was 16 Hz. The Symlet 5 wavelet was used, with a tree developed until level 3.

Results

Results on synthetic signals can be viewed on [3][17]. For uterine EMG ($F_e = 16 \quad Hz$), the training set consisted of a total of 100 signals containing 1000 events in total. Each signal was composed of 10 events of different types (contractions, Alvarez waves, LBDF waves or foetus motions) identified by an expert. The events were then contaminated by white noise (SNR: 10 db). The signals were decomposed by WPT. After applying the first step of the selection algorithm, only packets 1, 3, 7 and 8 (bandwidths: [0-4], [0-2], [0-1] and [1-2] Hz) were first selected (figure 5). They were easily selected according to the KS statistics by setting the threshold to 0.25 (figure 5). After applying all steps of selection of the best basis only packets 7 and 8 were retained.

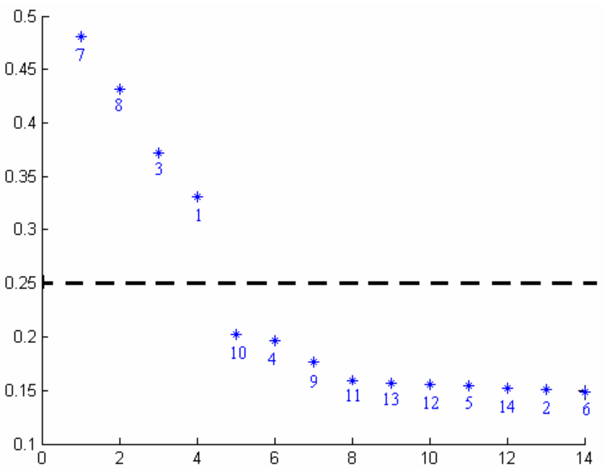

Figure 5: Average of KS statistics $D_{\max}$ from 100 uterine EMG signals on 14 packets. X axis: arbitrary unit. Y axis: $D_{\max}$ value.

To test the performance of the algorithm, 100 test signals were used with the same composition as for the training set. After correction of the change times [17] we obtained a detection probability of 0.9878 and a false alarm probability of 0.0545.

On the other hand, to demonstrate that the reduction of the number of packets by selection of the best basis did not change the above performance, the detection algorithm was applied on all wavelet packets selected before reduction (packets 1, 3, 7 and 8). The same 100 test signals were used. The new detection and false alarm probabilities were 0.9978 and 0.0652, respectively, i.e, we obtained roughly the same values of detection and false alarm probabilities, i.e. the useful information was preserved by the reduction process. Figure 6 illustrates the detection on a real uterine EMG signal.

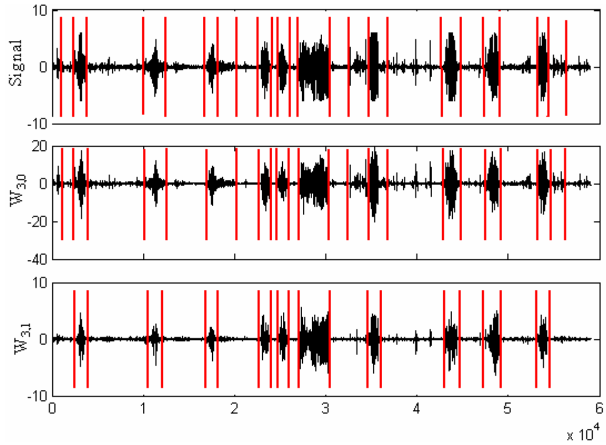

Figure 6: Detection algorithm applied to the selected packets. Vertical lines indicate the change times after correction. X axis: number of samples. Y axis: amplitude in arbitrary units.

Discussion

This work proposed the use of a WPT associated with a WP selection in order to define the best WP tree for event detection purposes. The choice of such a multiscale decomposition is justified by the fact that the recordings were characterized by their frequency contents. The algorithm of selection of best basis firstly led to the choice of a subset of WP based on a criterion of detection capability for a class of uterine EMG recordings, and then a reduction of the redundancy WP was processed. There wasn't any lose of information. The DCS algorithm was applied to the WP coefficients before and after the reduction of redundant WP. The results showed the same good detection and false alarm probabilities in both cases.

The criterion adapted to choose the best basis selection was the Kullback-Leibler Distance (KLD). When there was no event to be detected, in other words when there was only noise, the estimated KLD roughly followed an exponential distribution depending on only one parameter.

The fitting between sample histogram and exponential model was limited to equality between expectation and sample mean. Other ways could be obviously explored, for instance to keep the sample histogram obtained by simulation as a model. However the results produced by the proposed method again seemed to be highly acceptable when using the exponential distribution approximation.

ISSN: 1727-1983 © 2005 IFMBE

When events were detected in a packet, the distribution of the estimated KLD deviated from the exponential distribution. The statistics Kolmogorov-Smirnov Dmax was used to measure the separation between experimental and theoretical cumulative distributions in order to highlight the presence of ruptures, then to select the most relevant packets. Whatever the chosen statistics, the idea was to rank WP in descending order with respect to the value of the statistics, then to select only WP for which Dmax presented the highest values.

We asked an expert in uterine EMG to indicate, from arbitrarily selected examples, which WP enhanced the uterine events at best. She selected the same WP as the algorithm, in the same order. Without taking this expertise as universal evidence, this illustrated the concordance between an automatic unsupervised learning and a direct supervised selection.

Conclusions

The fact that the unsupervised method produced the same results as the supervised one makes it possible to achieve a selection process of the WPT packets without any need of either reference database or specific expertise for selection of the best basis. Event detection will be made only from the selected packets where the SNR is obviously highly improved. The use of a WP transform allows the easy addition of pre-processing steps like noise reduction. A further step would now be to associate a classification step based on the same WP decomposition, moving towards an identification of all detected events in the signal.

References

[1] KHALIL M., and DUCHENNE J. (2001): 'Uterine EMG Analysis: 'a Dynamic Approach for Change Detection and Classification', *Med.IEEE Trans. Biomed. Eng.*, 47, pp. 748-756.

[2] DEVEDEUX D., MARQUE C., MANSOUR S., GERMAIN G., and DUCHENE J. (1993): 'Uterine electromyography: a critical review', *Amer. J. Obstet Gynecol,* 169 (6), pp. 1636-1653.

[3] CHENDEB M., KHALIL M., and DUCHENE J. (2005): 'The Use of Wavelet Packets for Event Detection', Proc. Of EUSIPCO 2005- 13th European Signal Conf. Antalya, Turkey, 2005.

[4] WEI Z., YAM L. H., and CHENG L. (2004): 'Detection of Internal Delamination in Multi-Layer Composite Using Wavelet Packets combined with Modal Parameter Analysis', *Compos. Struct.,64*, pp. 377-387.

[5] PENG Z.K., TSE P. W., and CHU F. L. (2005): 'A Comparison Study of Improved Hilbert–Huang Transform and Wavelet Transform: Application to Fault Diagnosis for Rolling Bearing', *Mech. Sys. SignalProc.,* in press.

[6] COIFMAN R. R., and WICHERHAUSER M.V. (1992): 'Entropy based algorithms for best basis selection', *MIEEE Trans. Inform. Theory*, 38, pp. 713-718.

[7] SHIBATA R. (1997): 'Bootstrap estimate of Kullback-Leibler information for model selection', *Stat. Sin.,* **7**, pp. 375-394.

[8] SAITO N., and COIFMAN R. R. (1994): 'Local discriminant bases', *Mathematical Imaging: Wavelet Applications in Sig. and Image Proc. II, Proc. SPIE*, 2303, pp. 65-75

[9] RAVIER P., and AMBLARD P. O. (2001): 'Wavelet packets and de-noising based on higher-order-statistics for transient detection', *MSig. Proc.,* **81**, pp. 1909-1926.

[10] LEMAN H., and MARQUE C. (2001): 'Rejection of the Maternal Electrocardiogram in the Electrohysterogram Signal', *IEEE Trans. Biomed. Eng.*, **47**, pp. 1010-1017.

[11] HITTI E., and LUCAS M. F. (1998): 'Wavelet-packet basis selection for abrupt changes detection in multicomponent signals', Proc. of EUSIPCO 1998, 6th European Signal Proc. Conf., Island of Rhodes, Greece, 1998, p. 1841-1844.

[12] MALLAT S. (1999): 'A Wavelet Tour of Signal Processing', Academic Press, San Diego, CA.

[13] BONET J. S. D., and VIOLA P. (1998): 'Texture recognition using a nonparametric multi-scale statistical model', Proc. of IEEE Conf. Computer Vision Pattern Recognition, Santa Barbara, CA, 1998, p. 641-647.

[14] CHEN J. Y., BOUMAN C. A., and ALLEBACH J. P. (1997): 'Multiscale branch and bound algorithm image database, Proc. of IS & T/ SPIE Conf. on Storage and Retrieval for Image and Video Databases V San Jose, CA, 1997, p. 133–144.

[15] KULLBACK S., and LEIBLER R. A., (1951): 'On information and sufficiency, Ann. Math. Statis. pp. 79-86.

[16] DO M. N., and VETTERLI M. (2002): 'Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance', *Proc. IEEE Trans Image Process 11*., **2**, pp. 146-158.

[17] CHENDEB M., KHALIL M., and DUCHENE J. (2005): 'Selection of the best basis of wavelet packets tree and optimisation of detection algorithms', *Sig. Proc.,* in press.

ISSN: 1727-1983 © 2005 IFMBE