THE USE OF WAVELET PACKETS FOR EVENT DETECTION

Marwa Chendeb[1,2], Mohamad Khalil[2], Jacques Duchêne[1]

[1] ISTIT, University of Technology of Troyes,
UTT, 12 rue Marie Curie, BP 2060, 10010 Troyes, France
[2] Electronic Department, Faculty of Engineering, Lebanese University, Tripoli, Lebanon
Phone: + (33) 3 25 71 84 46, fax: + (33) 3 25 71 56 99, email: marwa.chendeb@utt.fr

ABSTRACT

In this paper, we propose a *best basis* selection method to choose a set of packets from a wavelet packet tree. Our goal is to obtain packets that show changes in both energy and frequency. The criterion adapted to choose the *best basis* is the Kullback-Leibler Distance (KLD). When there is no event to be detected, the estimated KLD follows roughly an exponential distribution depending on only one parameter: the length of the windows partitioning the signal. When events are detected in a packet, the distribution of the estimated KLD deviates from the exponential distribution. The statistics Kolmogorov-Smirnov are 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.

1. INTRODUCTION

There is a need to develop methods of detection and segmentation in real signals like Uterine EMG that is one of our application fields. Real signals are highly contaminated by noise resulting in a low signal to noise ratio (SNR). Therefore the extraction of relevant activities from real signals tends to be a complex task. Uterine EMG recordings are usually composed of activities with different frequency bands. When signals are decomposed using Wavelet Packets Transform (WPT), our problem is to choose from the WPT the packets that highlight the presence of ruptures. Then the detection algorithm on the selected packets will be applied.

Wei and colleagues presented a study on active detection using a combination of modal analysis and WPT [11]. Peng and colleagues used WPT and an effective method for intrinsic mode function (IMF) selection in the rolling bearing fault detection [6].
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[2], feature extraction and classification [10], etc. Ravier and Amblard presented a detector of transient acoustic signals combining the local wavelet analysis and higher-order statistical properties of the signals [8]. Leman and Marque proposed a more specific criterion to denoise the EHG signal [5]. Hitti and Lucas proposed a *best basis* selection method to detect abrupt changes in noisy multi-component signals [4]. They used energy criterion to allow separation of the different frequency components of the signal from a wavelet-packet library tree.

In our work, we use the wavelet decomposition WPT associated with a criterion of detection capability. If events and background activities mixed in the same recording present different probability densities, the Kullback Leibler distance KLD can be used as a distance between these densities. Therefore, KLD can be considered as an index of efficiency of each Wavelet Packet WP in detecting changes, hence events included in the recordings. For this reason, we use it as a criterion for choosing the most efficient or *best basis* for our application, i.e. the detection of different events that show both energy and frequency changes. KLD will be different according to whether there are one or more events in the packet. The idea is to approach the KLD statistics by a theoretical distribution where there is no event in a given wavelet packet WP. Hence a WP will be selected in the WP tree only if KLD significantly differs from that theoretical distribution. In addition, we apply the KLD directly on the wavelet packets' coefficients, rather than on the reconstructed signals.

2. METHODS

2.1 Wavelet Packet Transform (WPT)

WPT 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. 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. The sequence of functions is defined by:

$$W_{2n}(x) = \sqrt{2}\sum_{k=0}^{2N-1} h(k) W_n(2x-k)$$

$$W_{2n+1}(x) = \sqrt{2}\sum_{k=0}^{2N-1} g(k) W_n(2x-k)$$

where $W_0(x) = \phi(x)$ is the scaling function and $W_1(x) = \psi(x)$ is the wavelet function. In other words, the three indexed family of analyzing functions can be reached by:

$$W_{j,n,k}(x) = 2^{-j/2} W_n(2^{-j}x - k)$$

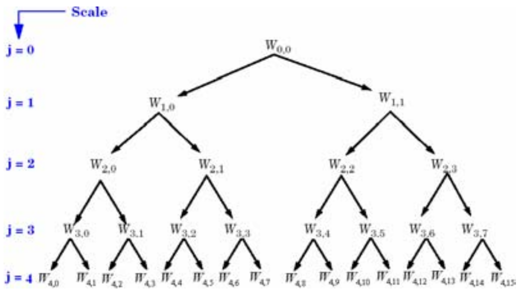

Figure 1: Wavelet Packet Decomposition Tree.

k can be interpreted as a time-localization parameter and j as a scale parameter. $W_{j,n,k}$ analyzes the fluctuations of the signal roughly around the position $2^j.k$, at the scale 2^j, and at various frequencies for the different admissible values of the last parameter n [1]. The wavelet packets' coefficients at each node *(j, n)* are written as:

$$\left(C_{jn}(k)\right)=\left\langle f(t),2^{-j/2}W_n(2^{-j}t-k)\right\rangle$$

2.2 Best basis selection

WPT allows a best adapted analysis of a signal. The idea is to select a suitable orthogonal subset basis from the general wavelet packets according to the objectives of the specific 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 [2].

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. In our work, the KLD is directly applied on a temporal partition of the packet coefficients, and not on the reconstructed signals.

2.3 Principles of the choice of the best basis

As previously mentioned, our goal is to find a set of nodes from the wavelet packet tree that allow detectability and classification of different events in the signal. To decide if a wavelet packet (or a node) contains more than one event, we make use of KLD as a criterion of presence of ruptures. Statistical properties of KLD estimated are related to the statistical characteristics of the WP coefficients. If we hypothesize that signal amplitude follows a Gaussian distribution, then a good PDF approximation for the marginal density of wavelet packet coefficients at a particular sub-band transform may be obtained by adaptively varying the two parameters of the generalized Gaussian density (GGD) [2], which is defined as:

$$f(x;\alpha,\beta)=\frac{\beta}{2\alpha\tau(1/\beta)}e^{-(|x|/\alpha)^{\beta}} \qquad \text{(a)}$$

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

2.3.1 Kullback-Leibler Distance

Estimation of the Kullback-Leibler Distance is a crucial part of deriving a statistical model selection procedure, and is based on the likelihood principle [8]. The KLD is considered as a measure of goodness of fit of a statistical model. Between two probability densities f_{θ_1} and f_{θ_2} of a random variable X, the KLD is defined by [2]:

$$D(f(X;\theta_1),(f(X,\theta_2))=\int f(x,\theta_1)\log\frac{f(x,\theta_1)}{f(x,\theta_2)}dx. \qquad \text{(b)}$$

Given the GGD, the PDF of packet coefficients in each sub-band can be completely defined by substituting (a) in (b). After some manipulations we obtain the following expression for the KLD between two PDFs of Gaussian models [2]:

$$D(f(.;\alpha_1,\beta_1);f(.;\alpha_2,\beta_2))=\log\left(\frac{\beta_1\alpha_2\tau(1/\beta_2)}{\beta_2\alpha_1\tau(1/\beta_1}\right)+$$

$$+\left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2}\frac{\left(\tau((\beta_2+1)/\beta_1\right)}{\tau(1/\beta_1)}-\frac{1}{\beta_1}$$

In our case, we assume that the original signal follows a Gaussian distribution. Consequently, the packet coefficients follow the same distribution as the WPT is a linear transformation. So, we fixed the parameter $\beta=2$.

If L is the length of the sequence x, an estimation of the parameter α is given as:

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

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

$$\hat{K}_{12}=\frac{1}{2}\left[\log\left(\frac{\hat{\alpha}_2}{\hat{\alpha}_1}\right)^2+\left(\frac{\hat{\alpha}_1}{\hat{\alpha}_2}\right)^2-1\right]$$

The KLD is not symmetric. To overcome this problem we use: $\hat{K}=\hat{K}_{12}+\hat{K}_{21}$

2.4 Distribution of the estimated KLD $\hat{K}$

Let us assume that the whole signal $X=(x_1,x_2,...,x_M)$ follows a Gaussian distribution with mean $\mu=0$ and standard deviation α , i.e. $\alpha_1=\alpha_2=\alpha$.

In each WP, X is partitioned into P consecutive segments of N points: $X_1=(x_{11},x_{12},...,x_{1N})$... $X_P=(x_{P1},x_{P2},...,x_{PN})$. For any pair of windows {l, m}, standard deviation is estimated as $\hat{\alpha}_l$ and $\hat{\alpha}_m$, respectively. Therefore, the estimated $\hat{K}_{lm}$ can be written as:

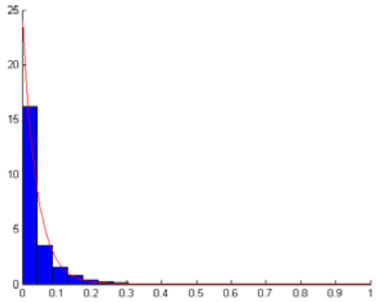

Figure 2: Histogram of $\hat{K}$ associated with the exponential distribution having the same expectation.

$$\hat{K}_{lm} = \frac{1}{2}\left[\log\frac{N.\hat{\alpha}_m^2}{\alpha^2} - \log\frac{N.\hat{\alpha}_l^2}{\alpha^2} + \left(\frac{\hat{\alpha}_l}{\hat{\alpha}_m}\right)^2 - 1\right]$$

For revealing the presence of ruptures among P segments, one may take $\binom{P}{2}$ pairwise combinations of $\hat{K}_{lm}$:

$$KLD = \sum_{l=1}^{P-1}\sum_{m=i+1}^{P}\hat{K}_{lm}$$

2.4.1 Expectation of KLD $\hat{K}$

$\frac{N.\hat{\alpha}_m^2}{\alpha^2}$ and $\frac{N.\hat{\alpha}_l^2}{\alpha^2}$ follow chi-square distributions with N degrees of freedom, hence $\frac{\left(\hat{\alpha}_l\right)^2/N}{\left(\hat{\alpha}_m\right)^2/N}$ follows a Fisher distribution. If X is a random variable following a chi-square law with N degrees of freedom, then the expectation of log X is given by this equation:

$$E[LogX] = (Log2 - \gamma) + \sum_{i=1}^{N/2-1}\frac{1}{i}$$

with $\gamma = Euler's$ Constant $\cong 0.5772$

The expectation of the estimated Kullback-Leibler distance $\hat{K}_{lm}$ is written as: $E\left(\hat{K}_{lm}\right) = \frac{1}{N-2}$

$$\text{thus, } E\left(\hat{K}_{lm} + \hat{K}_{ml}\right) = E\left(\hat{K}\right) = \frac{2}{N-2}.$$

2.4.2 Histograms of $\hat{K}$ *and first order approximation of the distribution*

The idea here is to approach the histogram of estimated values of $\hat{K}$ with a known distribution under the hypothesis $\alpha_l = \alpha_m = \alpha \quad \forall \quad l,m$. The empirical distribution of $\hat{K}$ is a function of only one parameter *N,* the degree of freedom, so that the theoretical distribution estimated by $\hat{K}$ histogram has to depend on one parameter. Taking into account the shape of the histogram of figure 2, we chose as a first approximation the exponential distribution, which is a special case of the gamma distribution, depending only on one parameter λ . Its probability density function is defined as:

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

In order to find λ , we equalize the expectation of $\hat{K}$ and that of the exponential distribution, i.e. $\frac{1}{\lambda}$:

$$\lambda = \frac{N-2}{2}.$$

Figure 2 shows an example of a histogram of $\hat{K}$. Note that the variance remains to be evaluated. The histogram shows that the selected approximation is valid only at order 1.

2.5 Wavelet Packets characterization

The goal of this study is to retain only the wavelet packets that contain at least one event. The characterization of a wavelet packet is that, if a wavelet packet contains at least one event, the distribution of $\hat{K}$ doesn't follow the exponential distribution. From the binary tree of wavelet packets, we put a value "0" or "1" at each node, according to the result of the comparison between the experimental $\hat{K}$ distribution and the corresponding exponential distribution (Fig 3a). Comparison is done by use of the Kolmogorov-Smirnov (K-S) test [7]. The K-S statistic $D_{\max}$ is the maximum difference between the theoretical cumulative PDF and the experimental distribution. For this issue we define the two hypotheses as:

H_0 : KLD follows exponential distribution.

H_1 : KLD does not follow the exponential distribution.

2.6 *Best basis* construction algorithm

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 strictly follows the first proposed by Hitti and Lucas [4]. 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.3a).
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.3b).
c. Only the nodes at "1" having a father at "2" or higher than "2" are selected in order to reduce the redundancy (Fig.3c).

Hitti's algorithm guarantees a complete basis representing the entire original signal (all the packets at 1 in Fig. 3c), i.e. the original signal can be reconstructed from the selected

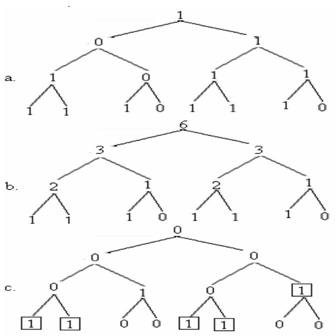

Figure 3: Steps for selection of the best basis.

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. 3c.

3. RESULTS

The algorithm of construction of best basis is applied on the wavelet packets obtained after decomposition of simulated signals. The simulated signals are composed of four segments of different frequency bands defined with respect to half the sample frequency: $F_e/2\ ([0.01;0.2],[0.21;0.4],[0.41;0.6],[0.61;0.8])$.

Segments follow a normal distribution with mean 0 and variance 2 with a signal to noise ratio equal to 10 dB. The instants of changes are 2000, 4000, and 6000. The signals length is 8000. The wavelet Symmlet 2 is used and the tree level number is 3. After basis selection, the CUSUM algorithm is applied on the coefficients of the selected packets [1].

Figure 4 highlights the detection result on a sample of a simulated signal. The selected nodes related to this signal are $W_{2,1}, W_{2,2}, W_{3,1}, W_{3,6}$ *and* $W_{3,7}$ (numbered in accordance with figure 1).

The best basis algorithm was applied on 100 simulated samples. The number of selected nodes was recorded for each of them. Figure 5 displays the selection rate of each node belonging to the best basis.

4. CONCLUSION

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 use of the KLD produced very satisfactory results from simulated signals. The method has now to be applied to real signals. Future works will be related to uterine EMG segmentation. A learning stage from recordings labelled by experts will define the best sub-tree. All further recordings will be then decomposed according to this tree before event detection on the selected wavelet packets.

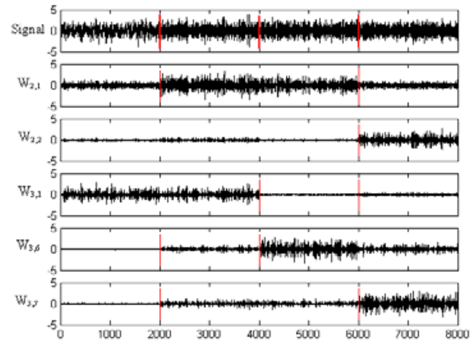

Figure 4: Detection algorithm is applied on 5 wavelet packets. Vertical lines show the segment detected by our algorithm.

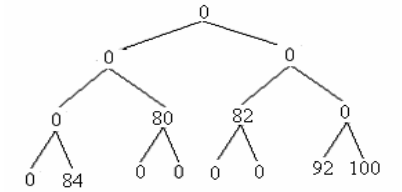

Figure 5: The best basis obtained in the third step of the construction of best basis (fig 3.c) is plotted. 100 test simulated signals are used in this simulation.

REFERENCES

[1] M. Chendeb, M. Khalil and J. Duchêne, "Wavelet Based Method for Detection: Application in Proprioceptive Rehabilitation", in *Proc 26th Int Conf IEEE EMBS,* San Francisco, CA, USA, September 1-5, 2004, pp. 161.

[2] R. R. Coifman and M. V. Wickerhauser, "Entropy based algorithms for best basis selection", *IEEE Trans Inform Theory*, vol. 38, pp.1241-1243, 1992.

[3] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance", *IEEE Trans Image Process*, vol. 11, pp. 146-158, 2002.

[4] E. Hitti and M. F. Lucas "Wavelet-packet basis selection for abrupt changes detection in multicomponent signals" in *Proc. EUSIPCO 1998*, Island of Rhodes, Greece, September 8-11, 1998, pp. 1841-1844.

[5] H. Leman, C. Marque, "Rejection of the Maternal Electrocardiogram in the Electrohysterogram Signal», *IEEE Trans Biomed Eng*, vol. 47, pp. 1010-1017, 2000.

[6] Z. K. Peng, P. W. Tse, F. L. Chu, "A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing", *Mech Syst Signal Proc*, in press, 2005.

[7] W. H. Press B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*. New York, USA: Cambridge University Press, 1988.

[8] P. Ravier and P.-O. Amblard, "Wavelet packets and de-noising based on higher-order-statistics for transient detection", *Signal Process*, vol. 81, pp. 1909-1926, 2001.

[9] N. Saito, R. R. Coifman, "Local discriminant bases". In A. F. Laine and M. A. Unser, Editors, *Wavelet Applications in Signal and Image Processing II*, vol. 2303, 1994.

[10] R. Shibata, "Bootstrap estimate of Kullback-Leibler information for model selection", *Stat Sin*, vol. 7, pp.375-394, 1997.

[11] Z. Wei, L. H. Yam, and L. Cheng, "Detection of internal delamination in multi-layer composites using wavelet packets combined with modal parameter analysis", *Compos Struct*, vol. 64, pp. 377–387, 2004.