31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Cardiac arrhythmia classification using Wavelets and Hidden Markov Models – A comparative approach

Pedro R. Gomes, Filomena O. Soares, J. H. Correia, C. S. Lima

Abstract — **This paper reports a comparative study of feature extraction methods regarding cardiac arrhythmia classification, using state of the art Hidden Markov Models. The types of beat being selected are normal (N), premature ventricular contraction (V) which is often precursor of ventricular arrhythmia, two of the most common class of supra-ventricular arrhythmia (S), named atrial fibrillation (AF), atrial flutter (AFL), and normal rhythm (N). The considered feature extraction methods are the standard linear segmentation and wavelet based feature extraction. The followed approach regarding wavelets was to observe simultaneously the signal at different scales, which means with different level of focus. Experimental results are obtained in real data from MIT-BIH Arrhythmia Database and show that wavelet transform outperforms the conventional standard linear segmentation.**

I. Introduction

The electrocardiogram (ECG) is the record of the electrical activity of the heart and provides fundamental information about its electrical instability being the most significant biosignal used by cardiologists for diagnostic purposes. Atrial fibrillation (AF) is perhaps the most common arrhythmia encountered in clinical practice, affecting about 0.5-1% of the general population. AF is not only related to frequent symptoms and reduced quality of life but also constitutes a major risk factor for stroke and mortality from cardiovascular and all other causes [1]. AF pathology is usually diagnosed based on ECG analysis.

Normally continuous monitoring over an extended period of time is required in order to increase the understanding of patients' cardiac abnormalities. Such situations require continuous monitoring by the physicians or alternatively the aid of automated arrhythmia detection equipment, which can be able to identify different types of arrhythmias.

This problem of cardiac arrhythmia detection can be viewed as a pattern recognition problem, since it is possible to identify a finite number of different patterns (arrhythmias).

Hidden Markov Models have been successfully applied to pattern recognition problems in applications spanning automatic speech recognition [2], image segmentation [3], ECG modeling [4] and cardiac arrhythmia analysis [5]. The most common approach regarding HMM training is finding the stochastic distribution that best fits the data. Usually this data is derived from the waveform from some type of signal processing usually known as feature extraction method. The most classical technique for feature extraction in the HMM framework is perhaps the linear segmentation where the ECG is segmented in straight line segments. More recently advanced signal processing techniques as Fourier Transform, Linear Predictive Analysis, Lyapunov Functions [6] and Multivariate Analysis (MA) have been used in order to overcome some limitations of the linear segmentation. Multivariate Analysis allows observing the signal at various scales emphasizing some hidden particularities not viewed at other scales. Wavelet Analysis is perhaps the most common form of multivariate analysis. Recently Wavelet Analysis was been successfully combined with Hidden Markov Models (HMMs) especially regarding ECG segmentation [11].

This paper reports the performance of two types of extraction feature methods evaluated under the conventional HMMs framework. The considered feature extraction methods are the classical linear segmentation [4], [7] where the ECG signal is linearized in order to discard some linear redundancy and the wavelet transform where the signal is simultaneously viewed at different scales.

The wavelet transform has the advantage over conventional techniques that time/frequency representation can be more accurately modeled by decomposing the signal in the corresponding scales. When the composition level decreases in the time domain it increases in the frequency domain providing zooming capabilities and instantaneous characterization of the signal [8]. This time/frequency representation which preserves both global and local information seems to be more adequate than linear segmentation for local characterization of the signal.

The baseline system is a Bakis or left-to-right Continuous Density Hidden Markov Models (CDHMMs) with a Gaussian Mixture Model (GMM) associated to each model transition. The ECG signal is previously sliced in singular pulses by using the Pan-Tompkins [9] algorithm and each pulse class is modeled by a six state model, modeling the Q-S, S-T, T, T-P, P and P-Q events.

Manuscript received April 7, 2009.

Pedro R. Gomes is with the Faculty of Engineering of University Lusiada, Largo Tinoco de Sousa, 4760-108 V. N. Famalicao Portugal, phone: +351-252-309-200; fax: +351-252-376-363; e-mail: pedroreis@fam.ulusiada.pt

Filomena O. Soares is with Industrial Electronics Department of University of Minho, Campus de Azurem, 4800-058 Guimaraes Portugal. e-mail: Filomena.Soares@dei.uminho.pt

J. H. Correia is with Industrial Electronics Department of University of Minho, Campus de Azurem, 4800-058 Guimaraes Portugal. e-mail: higino.correia@dei.uminho.pt

C. S. Lima is with Industrial Electronics Department of University of Minho, Campus de Azurem, 4800-058 Guimaraes Portugal. e-mail: carlos.lima@dei.uminho.pt

Experimental results from the MIT-BIH Arrhythmia Database using more than 1500 training pulses and 3000 testing pulses show that the wavelet transform is more adequate to extract information from de ECG signals than the standard linear segmentation procedure.

II. ECG Feature Extraction

ECG observations were obtained from the standard linear segmentation and wavelet based extraction methods.

A. Linear Segmentation

In the standard linear segmentation, observations were obtained from the segmentation of the original signal with straight line segments which goal is to decrease the amount of linear redundancy, as described in [4]. In [4] it is suggested for features a bi-dimensional vector where the components are respectively the amplitude of the starting point and the duration of the line segment. However, as reported in [7], these features are very sensitive to baseline wander, DC drift and heart rate variation. DC drift can be cancelled by using differential amplitude between the starting and ending points, and heart rate variability can be attenuated by normalizing the line segment duration by the R-R interval, as reported in [7]. The R-R interval is computed by using the Pan-Tompkins algorithm [9].

B. Wavelets

The wavelet transform (WT) is a signal representation in a scale-time space, where each scale represents a focus level of the signal and therefore can be seen as a result of a band-pass filtering.

Given a time-varying signal x(t), WTs are a set of coefficients that are inner products of the signal with a family of "wavelets" obtained from a standard function known as "mother wavelet". In Continuous Wavelet Transform (CWT) the wavelet corresponding to scale "s" and time location "τ" is given by

$$\psi_{\tau,s} = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right) \quad (1)$$

where ψ(t) is the mother wavelet, which can be viewed as a band-pass function. The term $\sqrt{|s|}$ ensures energy preservation. In the CWT the time-scale parameters vary continuously

$$\Psi_x^{\psi}(\tau,s) = \frac{1}{\sqrt{|s|}}\int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-\tau}{s}\right)dt \quad (2)$$

where the asterisk stands for complex conjugate. Equation (2) shows that the WT is the convolution between the signal and the wavelet function at scale "s". Therefore the shape of the mother wavelet seems to be important in order to emphasize some signal characteristics, however this topic is not explored in the ambit of the present work.

For implementation purposes both "s" and "τ" must be discretized. The most usual way to sample the time-scale plane is on a so-called "dyadic" grid, which means that sampled points in the time-scale plane are separated by a power of two.

As the scale represents the level of focus from the which the signal is viewed, which is related to the frequency range involved, then digital filter banks are appropriated to break the signal in different scales (bands). If the progression in the scale is "dyadic" the signal can be sequentially half-band high-pass and low-pass filtered.

The output of the high-pass filter represents the detail of the signal. The output of the low-pass filter represents the approximation of the signal, for each decomposition level, and will be decomposed in its detail and approximation components at the next decomposition level, and the process proceeds iteratively in a scheme known as wavelet decomposition tree, which is shown in figure 1. After the filtering half of the samples can be eliminated according to the Nyquist's rule, since the signal now has only half of the frequency.

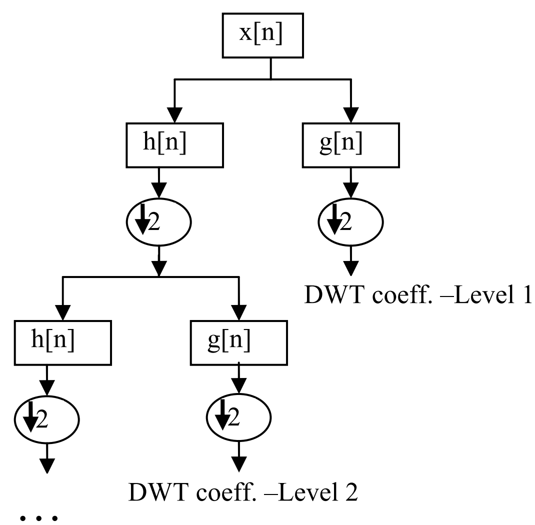

Figure 1. Wavelet decomposition tree.

This very practical filtering algorithm yields as Fast Wavelet Transform (FWT) and is known in the signal processing community as two-channel subband coder [10].

One important property of the DWT is the relationship between the impulse responses of the high-pass (g[n]) and low-pass (h[n]) filters, which are not independent of each other and they are related by

$$g[L-1-n] = (-1)^n h[n] \quad (3)$$

where L is the filter length in number of points. Since the two filters are odd index alternated reversed versions of each other they are known as Quadrature Mirror Filters (QMF). Perfect reconstruction requires, in principle, ideal half-band filtering. Although it is not possible to realize ideal filters, under certain conditions it is possible to find filters that provide perfect reconstruction. The most famous ones were developed by Ingrid Daubechies and they are known as Daubechies wavelets. In the ambit of this work only Daubechies wavelets with 2 vanishing moments (db-4) were used.

III. Wavelets Analysis of ECG

The multiresolution analysis based on the DWT can enhance small differences when the signal is simultaneously observed at the most appropriate scales. Figure 2 shows the result of the application of the DWT one cycle of a normal ECG. From the figure we can observe that d1 level (frequency ranges of 90-180Hz) emphasize the high frequency content of complex QRS when compared with P and T waves . D2 and d3 levels show clearly that other waves of small frequencies not seen at d1 scale are just appearing.

The features used in the scope of this work are simultaneous observations of d1 and d2 scales, therefore the observation sequence generated after the parameter extraction is of the form $\mathbf{O}=(\mathbf{o}_1, \mathbf{o}_2, \ldots \mathbf{o}_T)$ where T is the signal length in number of samples and each observation $\mathbf{o}_t$ is a bi-dimensional vector. Each element of the observation vector is derived from the IWT of the selected scale.

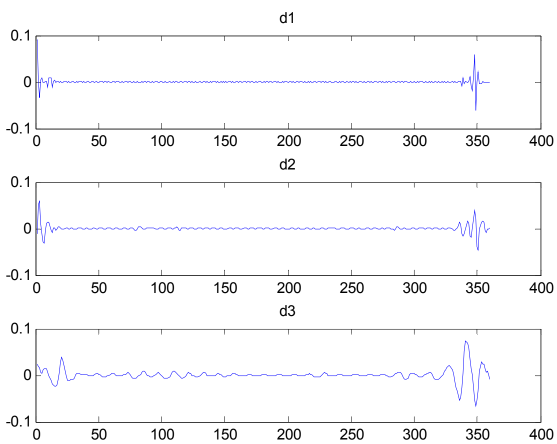

Figure 2. One ECG pulse viewed at scales d1, d2 and d3.

IV. Hidden Markov Models

Hidden Markov models are a doubly stochastic process in which the observed data are viewed as the result of having passed the hidden finite process (state sequence) through a function that produces the observed (second) process.

In an HMM the goal of the decoding or recognition process is to determine a sequence of hidden (unobservables) states (or transitions) that the observed signal has gone through. The second goal is to define the likelihood of observing that particular event, given a state sequence determined in the first process.

In the pattern recognition paradigm each class of beat is represented by a separate model and after decoding, the class for the which the probability (likelihood) of occurrence is greater is selected. Since the ECG is characterized by a time sequence waves occurring almost always in the same order which reflects the sequential activity of the cardiac conduction system an HMM structure where the states are connected in a left-to-right order was adopted. In [4] it is shown that a full connected HMM is eventually more appropriate for HMM modeling since the beat sequence reproduced by this kind of HMM is almost perfect. Figure 3 shows the model structure adopted for the several pathologies considered in the ambit of this paper.

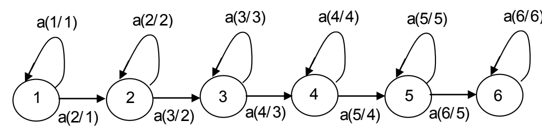

Figure 3. A left-to-right HMM with 6 states

The next issue is the choice of the number of Gaussian mixtures. For continuous models (CDHMMs), it has been found that it is more convenient and sometimes preferable to use diagonal covariance matrices with several mixtures, rather than fewer mixtures with full covariance matrices. The reason is the difficulty in performing reliable re-estimation of the off diagonal components of the covariance matrix from the necessarily limited training data. The HMMs in this work use five Gaussian mixtures per transition.

The output probability density function, which defines the conditional likelihood of observing a set of features when a transition through the model takes place, is usually a multivariate Gaussian mixture for the most engineering applications involving hidden Markov models. These probability density functions are associated with the transitions which configures a Continuous Density Hidden Markov Models (CDHMMs) Mealy machine and are given by

$$f(y/u_t) = \sum_{i=1}^{C} b_{u_t,i} G(y_t, \mu_{u_t,i}, \Sigma_{u_t,i}) \qquad (4)$$

Where c is the number of components in the Gaussian mixture, G(…) stands for bi-variate normal distribution with mean vector and covariance matrix for the i^{th} mixture component and transition u_t given respectively by $\mu_{u_t,i}$ and $\Sigma_{u_t,i}$. As the components of observation vector are assumed iid G(…) function in equation (4) is simply the product of five Gaussian functions. The mixture coefficients $b_{u_t,i}$ satisfy, for each transition u_t , to

$$\sum_{i=1}^{C} b_{u_t,i} = 1 \qquad (5)$$

so that, equation (4) is a probability density function.

In our experiments the observations were modeled by five components in the Gaussian mixture (C=5) in order to fit best data with multimodal distributions.

The Estimation of HMM parameters from a set of representative training data can be done by using the Baum-Welch algorithm which is based on the decoding of all the possible state sequence, or alternatively by using the Viterbi algorithm which is based on the most likely state sequence [2]. The adopted training was the MLE procedure in the Viterbi framework, which goal is to maximize iteratively the following probability density function

$$f(Y/\lambda) = f(Y/S,\lambda)P(S/\lambda) \quad (6)$$

where Y is the observation sequence, S the most likely state sequence and λ the set of HMM parameters. The model reestimation formulas can be found in [2]. This usual parameter estimation technique maximizes iteratively the model parameters that best fit the training data.

V. Experimental Results

Experimental results were evaluated by using the MIT-BIH Arrhythmia Database. Normal (N) and premature ventricular contraction (V) beats, in atrial fibrillation (AF), atrial flutter (AFL) and normal (N) rhythms were selected.

The training set contains the 121, 122, 221 and 222 records and the testing set contains the 105, 112, 121, 122, 210, 221 and 222 records of the MIT-BIH arrhythmia database. For the training set 722 normal (N) pulses of 121 (N rhythm) and 122 (N rhythm), 682 normal and premature ventricular contraction (V) pulses of 221 (AF rhythm) and 197 normal pulses of 222 (AFL rhythm) records were used. The testing set contains 2065 pulses of 105, 112, 121 and 122 records, 1011 pulses of 210 and 221 records and 246 pulses of 222 record, which means that data for training and testing purposes was obtained from different patients, which is normally known as patient-independent analysis. Tables 1 and 2 show respectively the HMM based pulse classification system using respectively features from linear segmentation and wavelets.

Table 1 – The confusion matrix associated to linear segmentation

	AF N	AF V	AFL N	N N	Total	Pr+
AF N	868	0	0	335	1203	0.72
AF V	0	114	0	0	114	1
AFL N	7	0	246	13	266	0.92
N N	22	0	0	1717	1739	0.98
Total	897	114	246	2065	3322	
Sensitivity	0.96	1	1	0.83		

Both feature extraction methods use the same dimensionality since only two first scales of the DWT were used in order to increase the accuracy regarding comparative performance. The row labeled "Total" means the total number of beats used in experiment for each class listed in the corresponding column. Both MLII and V1 signals were used each one with their own HMM. A pulse is considered classified if the score from both models agree, otherwise the pulse is considered wrong. Regarding the linear segmentation algorithm various initial parameters were tried, however the best performance was obtained wish $\varepsilon = 0.01$, max = 1 and k = 2 [4].

Table 2 – The confusion matrix associated DWT

	AF N	AF V	AFL N	N N	Total	Pr+
AF N	864	0	0	0	864	1
AF V	0	114	0	0	114	1
AFL N	0	0	237	0	237	1
N N	33	0	9	2065	2107	0.98
Total	897	114	246	2065	3322	
Sensitivity	0.96	1	0.96	1		

Comparing table 1 and table 2 it is clear that wavelet transform leads to a better characterization of ECG records, at least concerning HMM classification.

VI. Conclusion

The main conclusion is that the wavelet transform outperforms the linear segmentation regarding beat classification. In fact Multivariate Analysis seems to have higher potential than linear segmentation regarding spectral content signal analysis especially in relative quiet environments.

References

[1] V. Fuster, L. E. Ryden, R. W. Asinger, et al."ACC/AHA/ESC guidelines for the management of patients with atrial fibrillation: Executive summary," *Circulation*, vol. 104, pp. 2118-2050, Oct. 2001.

[2] Rabiner, L. R. (1989). *A tutorial on hidden Markov models and selected applications in speech recognition*. Proc. IEEE vol. 77, pg. 257-286.

[3] Choi, H. and Baraniuk, R. G. (2001). *Multiscale image segmentation using wavelet-domain hidden Markov models*. IEEE Trans. Image Process., vol. 10, no. 9, pp 1309-1321.

[4] Koski, A., (1996). *Modelling ECG signals with hidden Markov models. Artificial Intelligence in Medicine*, Vol. 8, pp. 453-471.

[5] B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.

[6] A. Casaleggio, S. Braiotta, A. Corona (1995). *Study of the Lyapunov Exponents of ECG Signals from MIT-BIH Database.* Computers in Cardiology.

[7] Cheng, W. T. and Chan, K. L. (1998). *Classification of Electrocardiogram using hidden Markov models*. Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 20, No. 1, pp. 143-146.

[8] Mallat, S. (1988), A wavelet tour of signal processing, Academic Press.

[9] Tompkings WJ (1995). *Biomedical Digital Signal Processing*. Prentice-Hall, Upper saddle River, NJ.

[10] Rangayyan, R. M. and Lehner, R. J. (1988). *Phonocardiogram signal processing: A review*. CRC Critical Reviews in Biomedical Engineering, 153 (3): 211-236.

[11] Andreão, R.V., Dorizzi, B. and Boudy, J. (2006). *ECG Analysis using Hidden Markov Models.* IEEE Transaction on Biomedical Engineering, 53(8):1541-1549.