DE-NOISING AND STATISTICAL FEATURE EXTRACTION OF THE ECG SIGNAL USING WAVELET ANALYSIS

TANYA KAMBO, RAM AVTAR JASWAL

Abstract—The electrocardiogram is a technique of recording bioelectric currents generated by the heart which is useful for diagnosing many cardiac diseases. The feature extraction and de-noising of ECG are highly useful in cardiology. And baseline wander elimination is considered as a traditional problem. So this paper deals with the noise removal and feature extraction of the signal. This baseline wander noise is removed by Butterworth filter. We present a wavelet-transform (WT) based search algorithm. The algorithm computes wavelet packet coefficients and then in each scale the different nine features of the signal is calculated. Comparison is made between the characteristics of signals and the branch of the wavelet binary tree corresponding to minimum entropy wavelet spaces is chosen. This algorithm is tested using the data record from MIT/BIH database for arrhythmia disease categorized into normal and abnormal and excellent results are obtained.

Keywords—ECG signal, de-noising, feature extraction, MIT database, wavelet packet decomposition.

I. INTRODUCTION

Electrocardiogram (ECG) is nearly a periodic signal that imitates the activity of the heart. Much information on the normal and abnormal condition of heart can be obtained from ECG. As the ECG signal being non-stationary in nature, it is very difficult to examine the signal visually. Thus the need arises for computer based methods for ECG signal analysis. Observation of ECG signal in hospitals can take long hours and can result in tiredness. And so we cannot be relied upon visual analysis. This calls need for computer-based techniques for signal analysis. At every beat, the heart is depolarized to trigger its contraction [1]. This electrical activity is transmitted throughout the body and can be picked up on the skin which is the principle behind the ECG. An ECG machine records this activity via electrodes on the skin and displays it graphically on the screen. An ECG involves attaching of 10 electrical cables to the body: one to each limb and six across the chest [6].

ECG is a wave that represents an electrical event in the heart such as atria depolarization, ventricular depolarization and repolarization. One cardiac cycle of ECG signal consists of the P-QRS-T wave. The major useful information in the ECG is taken from the intervals and amplitudes called as its features. As for the analysis of the signal it must be noise free [5]. As when we record the signal it contains many noises like power line interference and baseline wandering. So to removethese by using appropriate digital filters like fir filter and Butterworth filter [9,11]. There are different type of noise is present in the signal such as powerline interference, motion artifacts and baseline wandering. Motion artifacts are due to motion of the patient while recording the signal [11]. These can be filtered using different digital filters like FIR and IIR filters [11] but here we have chosen Butterworth filter due to its simplicity. And the main consideration is of removal of baseline wander. And the estimated this noise can detect the diseases [2]. And then extracting of features is an important tool for the analysis. There are several methods to determine non-linear i.e. statistical features like wavelet transform, FFT, STFT [5]. But we have first decomposed the signal using wavelet packet decomposition [4]. These coefficients are used for the non-statistical features [5, 8] which tells about the original features of the signal. The implementation of wavelet decomposition is easy as compared to and it generates the wavelet tree [4]. So after having coefficients and extracted features we can differentiate the signals according to their diseases [2]. Biomedical research in signal processing of these biomedical signals are in very trend [4]. These are very efficient and interesting to read and study.

II. DATA SELECTION

The collection of data is the one of the most important chore of signal processing. In this work, MIT–BIH Arrhythmia database directory of ECG signals from physio.net ATM BANK is used. The ECG signal of MIT–BIH Arrhythmia was obtained from the Beth Israel Hospital Arrhythmia Laboratory [8]. This database contains 48 files partitioned into two parts first one is of 23 files (numbered from 100 to 124 inclusive with some numbers missing) selected at random from this set, and another one contains 25 files. Each of the 48 records is slightly over 30 min long [2]. The ECG signals from MIT–BIH database are illustrated by – a text header file (.hea). The header files describes the detailed information about the signal such as number of samples, sampling frequency and interval, format of ECG signal, type and number of ECG leads. The sampling frequency of the signal is 360Hz. Two records have been taken 105 and 118 from this database. These records have analyzed to do comparison between them in their
features. According to [2] one is taken as normal and another is abnormal.

III. PREPROCESSING

This process is the first and very vital stage of signal processing, in which it is necessary to eliminate noises from the signal by using appropriate filters. For pre-processing of the ECG signal, noise elimination comprises various approaches for many noises [11]. This pre-processing of ECG signal is ended before extracting the features, which results in accurate features to increase the efficiency of the system. Preprocessing of ECG signal contains of de-noising of ECG signal and baseline wander removal using Butterworth filter.

3.1. Baseline wander removal

This type of noise artifacts that generally affects ECG signals is baseline wandering. Generally it appears from respiration and deets between 0.15Hz and 0.3Hz [3]. It is necessary to eliminate this noise in order to do ECG signal processing and to remove the irregularities from the signal. Electrode and respiration factors varies due to patient are vital sources of this baseline wander in most of the ECGs recordings. In this, the baseline wander of ECG is removed by first packing the signal taken from database then smoothing the data by using Butterworth filter of order 10[9]. This is a digital filter which has good efficiency and simple to use. Also it gives better result. It generates the coefficients known as Butterworth coefficients which helps in removing the noise [7].

3.2 Wavelet packet decomposition

Wavelet packet analysis is used to determine features of the non-stationary like ECG signal and this is the best method for finding features from ECG signal as it decomposed the signal at various level in different frequency bands. Wavelet transform gives time-frequency representation of a signal and it gives better resolution [7]. Wavelet packet decomposition analysis is a generator effective form of the DWT, in which signal is riven into approximation and detailed coefficients. The approximation coefficients is again riven up into a second level approximation and detailed coefficients, and this process is repeated till we want get result. There are k+1 possible ways to encrypt the signal for the k-level decomposition [3]. The signal is de-noised by low and high pass filter called quadrature mirror filters. In wavelet packet analysis approximation and detailed coefficients can be formed like a complete binary tree structure. This decomposition process applies at both low and high frequency sub – bands to create next level in the binary tree, it produces more than 2^{k-1} different ways to encode the signal. For each level k, there are 2^{k-1} number of nodes remained after wavelet packet decomposition [3]. Fig 1 shows the wavelet packet decomposition tree obtained at the 4th level of the decomposition and data for node (4,15) means data at 4th level and 15th coefficient of decomposition. Shannon entropy based criterion is used to find this feature. Fourth level decomposition coefficients are used to extract the features.

IV. FEATURE EXTRACTION

The pre-processed ECG signal is highly non-linear and requires statistical feature extraction methods to extract the features from the ECG signal. The Short Time Fourier Transform (STFT) and Fast Fourier Transform (FFT) has been used by numerous researchers for ECG signal processing. However, there are some limitations are still existing in FFT due to its dearth of response of time information's while analyzing it in frequency domain and vice versa [5]. In order to come over this limitation, the time frequency domain (TFD) analysis is required, which is carried out by using the discrete wavelet transform, empirical mode decomposition (EMD), Smoothed Pseudo Wigner-Ville Distribution (SPWVD) and etc. are used to derive the time and frequency domain features. In this, we have used DWT with “db4” one of the wavelet function which calculates statistical features. Because the wavelet characteristics of “db4” looks like ECG signal and it has been used by some researchers on ECG signal processing and for de-noising [10]. In order to extract features, some statistical properties are used.

4.1 Features

As signal is decomposed to determine its features. Statistical and non-statistical properties such as energy, mean, median, entropy, standard deviation, skewness, kurtosis, covariance is calculated from the coefficients which is generated after decomposition to create the feature set [4]. All algorithm are developed in the MATLAB environment with wavelet. The coefficients C_{ik} at 4th level of decomposition are used to calculate the features by following standard equations;

Energy, \( E = \sum_{k=1}^{n} C_{ik}^2 \)  

Standard deviation, \( \sigma_i = \frac{1}{n-1} \sum_{k=1}^{n} (C_{ik} - \mu_i)^2 \)  

Mean \( \mu_i = \frac{1}{N} \sum_{k=1}^{n} C_{ik} \)  

Kurtosis, \( \text{KUR} = \frac{\frac{1}{n} \sum_{k=1}^{n} C_{ik} - \mu_i^4}{\sigma_i^4} \)  

Skewness, \( \text{SK} = \frac{\frac{1}{n} \sum_{k=1}^{n} C_{ik} - \mu_i^3}{\sigma_i^3} \)  

Entropy, \( \text{EN} = \sum_{k=1}^{n} e^{2 C_{ik}} \log (C_{ik}^2) \)  

Median, \( \text{MED} = \text{median} \)  

Covariance, \( \text{COV} = E[(C_{ik} - \mu_i) \ldots (C_{i,j} - \mu_j)] \)  

Variance, \( \text{VAR} = (\sigma_i^2) \)  

The coefficients C_{ik} is the decomposition coefficient. Here i=1,2,…,l is the node number at 4th level of decomposition. N is the number of coefficients at
each decomposed level [4]. Energy, standard deviation, entropy, median are computed for each node at 4th level of decomposition tree. Total nine properties, as defined in above equations, are used for creating the feature set. Thus at 4th level of decomposition, there are 16 coefficients (2^4 = 16) coefficients. This is called wavelet tree.

![Wavelet tree](image)

**Fig1. Wavelet tree**

V. RESULTS

![Signal characteristics of record 105](image)

**Figure2: signal of record 105 (normal) database and its characteristics**

![Wavelet packet decomposition](image)

![Statistical plots](image)
CONCLUSION

In this paper, we have extracted nine features of the signal which determines the characteristics of the signal. These features tells the difference between the two signals. As these features gives information which helps us to find the condition of the signal i.e normal or abnormal. These features have the different values in both the signals which tells us about the condition. So these non-statistical have great significance in this biomedical research.

FUTURE SCOPE

The data is taken from the MIT database for analysis. So after extracting the features of the signal they can be classified using classifier and accuracy and sensitivity can be obtained using the features of these signals.

REFERENCES


