Improved of PRD for ECG signal decompression

CARLOS M. TRAVIESO, JESÚS B. ALONSO, MIGUEL A. FERRER.
Departamento de Señales y Comunicaciones
Universidad de Las Palmas de Gran Canaria
Campus de Tafira, E35017, Las Palmas de Gran Canaria. SPAIN

URL: http://www.gpds.ulpgc.es

Abstract: - This present work develops a decompression algorithm based on minimum mean-square error (MMSE), for improving the quality of ECG signal after its compression. This system has been made for improving the compression FAN algorithm and using the MIT Arrhythmias Database, because this database has ECG signals with and without pathologies, more similar to real condition. The results show Percent Root-mean-squared Difference (PRD) is three times better than the classical decompression.

Key-Words: - Decompression, ECG signal, Polynomial regression, MMSE, Interpolation, PRD, FAN algorithm.

1 Introduction

Nowadays, the technological advances are been applied each time more to the medicine. At present, there is a large quantity of signals that are generated and they are saved at the hospitals. These technological advances permit the data acquisition and storage is digital way, in such a way that the patient medical record can be seen by any communications network.

Particularly, one of the most used signals is the Electrocardiogram (ECG). Its use is very useful for medical doctors, due to establish diagnoses and to carry out monitoring of the patients.

The electrocardiogram (ECG) is a time-varying signal that reflects the electric current that stimulates cardiac muscle to contract and relax [1][2]. The ECG signal is acquired using at least two electric sensors applied at specific points of the patient's body, recording the electric potential between these two points. Basically, heart activity can be described as a continuous two step movement, depolarization/repolarisation, contraction and relax of the muscle. A single normal cycle of the ECG represents the atrial depolarization/repolarisation and the ventricular depolarization/repolarisation present in every beat. Each of the above mentioned movements are reflected in the ECG waveform and labeled as P, Q, R, S, and T waves [1][2]. All these waves are fundamental for the cardiologists, since they extract the information for their diagnoses.

One of the most common values that are obtained from the extraction is the RR' interval, or its inverse, the beep rate. Heart Rate Variability (HRV) can show the presence of abnormal heart activity. Several other time intervals, such as the QT interval or the ST segment, can also determine pathologic beats. Not only are time related values extracted from the ECG signal, but also amplitude related values like the QRS complex area or even frequency related values, such as LF/HF ratio, defined as the power difference between 0.015 Hz and 0.15 Hz, are obtained. Using all these values previously extracted as an input to a classifying system can provide the cardiology specialist with an extremely useful tool capable of automatically detecting the presence of abnormal beats in a patient's ECG [1][2].

Cardiology departments have become one of the most important areas within Health Services all over the world, due to the number of deaths caused by cardiology related problems. In Spain, according to the National Statistics Institute (INE), approximately a third of the total number of fatalities had its origin in a cardiac disease. As the single largest cause of death it cannot be ignored.

Nevertheless, research carried out in this area to date is considered to be insignificant, bearing in mind the high rate of deaths caused by cardiac diseases. European Union countries invest on average a minimal value of 3%, considerably less than countries such as the USA (6.8%) or South Korea (5.4%).

A fundamental aspect is the storage of this signal. Due to great number of samples stored, the research should be centred in the compression of the data. There are two methods of compression, loss and lossless compression systems. The lossless compression systems are very useful for the medical doctors since they have all information, but the compression rate is smaller than lossless data compression. The latter obtains better compression, but the medical doctors lose information from the signal. For this reason in this paper, we have worked in the improvement of the decompression signal on a loss compression system. We have developed a non-linear interpolator system, which improves the quality of the ECG signal, since all the waves are important for a medical diagnosis.

In the following figure, it can be observed the blocks diagram of this paper,

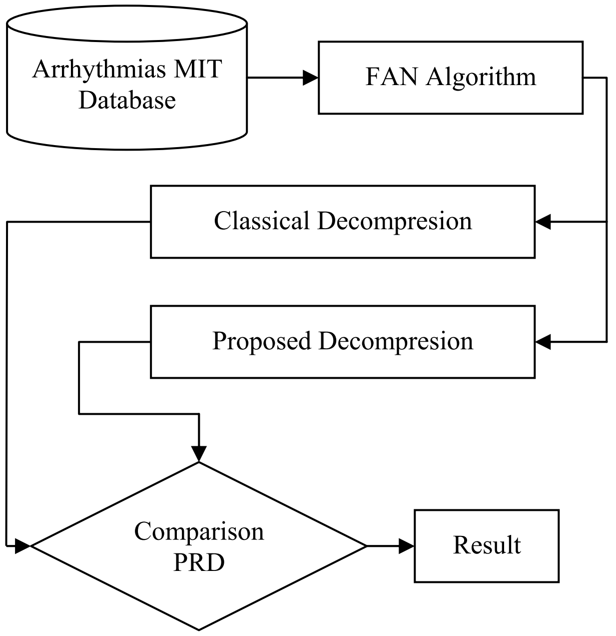

Figure 1. Proposed Blocks Diagram

The used database is MIT Arrhythmia Database [3], which presents a large quantity of pathological and healthy data. Therefore, this paper works with real conditions.

The used decompression technique is by minimum mean-square error (MMSE), which calculates the best function for the decompression, versus a classical lineal interpolation.

Finally, to comment that the evaluation of the system will be carried out by means of the parameter Percent Root-mean-squared Difference (PRD), which is one of the most used methods of quality in the bibliography [1][2].

The remainder of the paper is divided into the following sections. In the Section 2, the used database is described. The section 3 contains the algorithm of compression and decompression utilized in this work. In the section 4, the used improvement algorithm is commented. Experiments and results are shown the section 5, and subsequently, in the section 6 the conclusions are presented. This paper finalizes with references.

2 MIT Arrhythmia Database

The MIT Arrhythmia Database will be used for the classification system testing process. This database is composed of 48 half hour length records, sampled 360 samples per second with 11 bits resolution on a range of 10 mV. These 48 records are divided into two groups: 23 with common arrhythmias, and the remainder, a further 25, presenting abnormal cases. All these records have been labeled to facilitate their reading and for testing purposes. The MITDB is the most commonly used, and is generally regarded as the standard in ECG classification.

In this work, a group of 23 records presenting common arrhythmias has been selected in order to check the performance in usual scenarios.

It can be appreciated the database contains a great variety of pathological signals. This does it really useful for the implementation of the compression and the decompression. Therefore, the diversity of the present signals in the database presents a large number of real conditions, with healthy and pathological signals.

3 Compression algorithm

About medical signals compression, and in particular ECG signals, our goal for the data compression:

- Increase the storage capacity of the ECG signals databases. These data are habitually used for the study and classification of ECG signals, and they are composed by a large quantity of registrations.
- Increase the data broadcast that have been obtained or they are being acquired (real time) through a communication channel. If this way is a normal telephone line, then the compression becomes an indispensable tool.
- Enlarge the functionality of the screens or monitors, and storage systems of outpatient departments and hospitals.

The main objective of the data compression is going to obtain the greater data reduction, preserving the useful characteristics of the signal. Under this point of view, it will be possible, on some determined cases, to eliminate samples from the ECG signal; these do not contribute with the useful information for the cardiologist.

In the compression by parameters extraction (as FAN algorithm), our compression system extracts from ECG signal some characteristic parameters, as for example, amplitudes and duration times from standard segments, and these are stored and/or transmitted instead of the ECG signal, therefore, the original signal is lost. By this reason, this type of compression for ECG signal is not accepted among cardiologists, since a medical point of view is preferably to do observation and diagnosis on the original signal.

3.1 Compression FAN algorithm

The FAN algorithm [1] is based on the polynomial interpolator of first degree with two degrees of liberty (FOI-2FD), and is method, that better results offers in ECG compression.

Its operation begins accepting and storing the first sample X_0. After collecting the second sample, X_1, it is generated two straights, U_1 and L_1 (see figure 2). The first straight draws from X_0 to $X_1 + \varepsilon$, where ε represents the tolerance of the algorithm, and it gives quantity/quality relation of the obtained compression. The second straight draws from X_0 to $X_1 - \varepsilon$. If the third sample (X_2) is located in the area between the previous two straight, then, it is generated two new straights (U_2 and L_2), which will draw from points X_0 to $X_2 + \varepsilon$, and from X_0 to $X_2 - \varepsilon$, respectively. After this, between four drawn straights (U_1, L_1, U_2, L_2), we have chosen the two straights of them, which have the more restrictive area. The following step, it is changed the value of X_2 for the value of X_1, and a new sample is collected. This process is repeated until the new sample is located out of the area between straights. When this occurs, it is saved the length of the line (Time index), and its final amplitude (X_1). The sample X_2, which remains out of the area, is taken like X_0 and the algorithm begins since its beginning. In the figure 2 is shown this process of a graphic way. Like it is obtained the straights U and L from samples taken from n_1 to n_2, and which samples are eliminated and which stored. Too, this figure shows the process from U_1 and L_1 to U_2, L_2 and X_0.

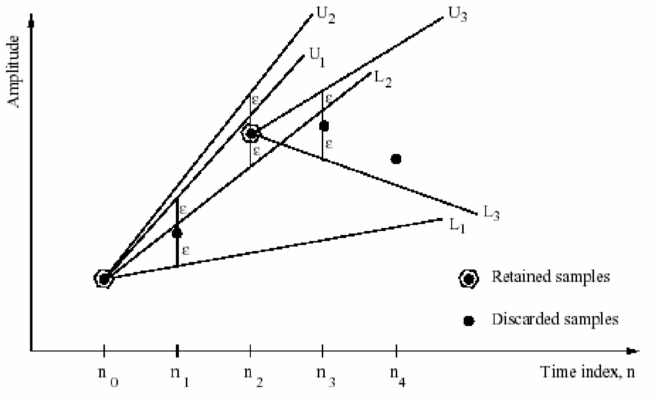

Figure 2. Fundamentals of FAN algorithm.

3.2 Percent Root-mean-squared Difference (PRD)

An instrument of measure that link the original signal with the decompression signal, in such of that it is found a agreement between the factor of reduction and the quality of ECG signal, so that the cardiologist can use that information. In the following expression, it can be observed like the percentage of the mean-quadratic difference, which is a measure of the quality of the error; the difference among original and decompression signals.

$$PRD = \sqrt{\frac{\sum_{i=1}^{N}\left(x_{org}(i) - x_{rec}(i)\right)^2}{\sum_{i=1}^{N}\left(x_{org}(i)\right)^2}} \cdot 100\% \qquad (1)$$

where;

- N is the number of samples
- X_{org} are samples of the original signal
- X_{rec} are samples of the decompression signal

4 ECG Decompression by polynomial regression

On the application of the neural networks [4], an important goal is to provide the introduction of a simple way, an appropriate polynomial curves given an set of N points from the technique of the minimum mean-square error.

Considering a polynomial of degree M,

$$\begin{aligned} y(x) &= w_0 + w_1 \cdot x + w_2 \cdot x^2 + \ldots + w_M \cdot x^M \\ &= \sum_{j=0}^{M} w_j \cdot x^j \end{aligned} \qquad (2)$$

This is a no-lineal reconstruction, which takes "*x*" as input, "*y*" like output.

The function *y(x)* is determined for parameters w_0, ..., w_M, that is analogous to weights on a neural network ($w = w_0 \ldots w_M$). Therefore, the polynomial can be written like $y=y(x;w)$.

Data have as index n=1,..., N, (nomenclature x^n), corresponding to desired value for the input. For the output is given by t^n.

To find values of the polynomial coefficients (*w*) is convenient to consider the error between the desired output (t^n), for a private entrance x_n); and the corresponding predicted value by the polinomial function ($y=y(x^n;w)$).

This procedure of polynomial curve fitting is calculated minimizing the error, to add all the data;

$$E = \frac{1}{2}\sum_{n=1}^{N}\left\{y\left(x^n;w\right) - t^n\right\}^2 \qquad (3)$$

As this function depends on w, we can minimize it. Therefore, the polynomial *y(x)* is a lineal function of w; but the error (*E*) is a quadratic function of *w*. For the input data, this corresponds to no-lineal function.

This minimum mean-square error, which uses the values at the desired output (t^n) as the function of the output, it is called learning supervised.

From the error expression, the approximation polynomial function is proposed, through the minimization of the error, to determine the polynomial coefficients that minimize it.

$$\nabla_w E = 0 \quad \rightarrow \quad \frac{dE}{dw} = 0$$
$$\frac{2}{2}\sum_{n=1}^{N}\{y(x^n, w) - t^n\}\cdot\sum_{n=1}^{N} x^n \tag{4}$$

since *E* depends lineally of *w*, and passing to matrix nomenclature, the development of the “derivate equal to 0” is,

$$\begin{aligned} &[(x_{N\times M}\cdot w_{M\times 1}) - t_{N\times 1}]\cdot x_{N\times M} = 0 \\ &x_{N\times M}\cdot w_{M\times 1} - t_{N\times 1} = 0 \\ &x_{N\times M}\cdot w_{M\times 1} = t_{N\times 1} \\ &w = x^{-1}\cdot t = t/x \end{aligned} \tag{5}$$

5 Experiments and results

We have developed experiments with the MIT Arrhythmias Database [3]. This methodology used the result of application of the compression FAN algorithm, and after, applied the decompression system by the classical technique of lineal reconstruction, and too, by the system proposed of regression polynomial.

Minimum mean-square error was calculated from degree of 2 to 5, and the variation of ε was done with the interval [0.02,∞). Results achieved of the PRD with the previous variables can be observed in the table 1, from the lineal method to the proposals.

About the observation of the results, we can be observed always an improvement on the no-lineal system presented, where the best result is obtained for a polynomial of degree 2. And where the resulting improvement, it is three times better than the value of the initial PRD. It presents a very significant improvement for the analysis and monitoring of ECG resultant. For values of ε greater to 0.002, the algorithm always returns the same values, since with a smaller one there is not significant variations for the algorithm.

In the figure 3, we have shown the comparison of the decompression (blue line) and original (green line) signals. The difference between lines is very small.

ε	Lineal	Degree of polynomial			
		2	3	4	5
0.02	8.89%	3.27%	5.16%	6.35%	22.52%
0.01	6.47%	2.42%	3.74%	3.78%	6.86%
0.009	4.78%	1.78%	2.58%	2.78%	3.71%
0.007	4.27%	1.54%	2.68%	2.95%	3.36%
0.005	4.20%	1.52%	2.66%	2.94%	3.24%
0.003	2.34%	0.81%	1.39%	2.76%	1.53%
≥ 0.002	1.42%	0.46%	1.01%	2.53%	0.90%

Table 1. Results of PRD for decompression

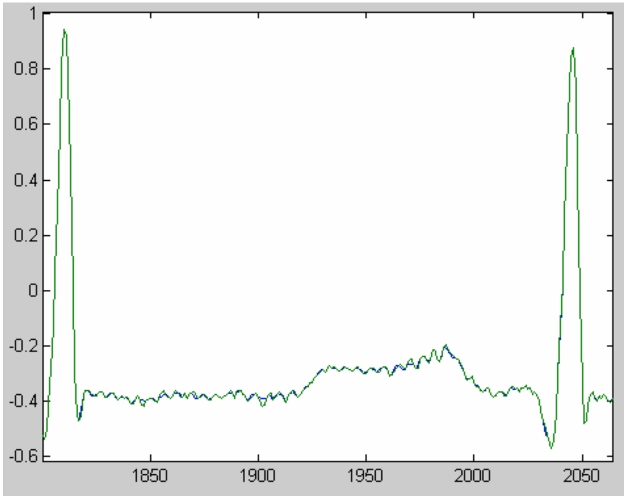

Figure 3. Original and decompression pathological ECG signal (degree 2 and ε = 0.003).

6 Conclusions

On this present work, we have made a robust and improved algorithm for ECG decompression, on the FAN compression algorithm. We have found on all our experiments this algorithm gives better results (PRD) versus the typical decompression system. Besides, these experiments have been made with the MIT Arrhythmias Database; this is, for real condition of pathological and healthy ECG signal. These results improve three times the quality of the decompression ECG signal.

References:

[1] W.J. Tompkins, *Biomedical Digital Signal Processing*, Prentice Hall, 1992.

[2] J. Bronzino, *The Biomedical Engineering: Handbook*, CRC Press, Springer, IEEE Press. 2000.

[3] http://www.physionet.org/physiobank/mitdb. *MIT-BIH Arrhythmia Database.* Last Visit: 22-10-2006.

[4] C.M. Bishop, *Neural Networks for Pattern Recognition,* Oxford University Press, 1997