Computers and Electrical Engineering 56 (2016) 30–45

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

# A 2D electrocardiogram data compression method using a sample entropy-based complexity sorting approach☆

Anukul Pandey[a,*], Barjinder Singh Saini[a], Butta Singh[b], Neetu Sood[a]

[a] Department of Electronics and Communication Engineering, Dr B R Ambedkar National Institute of Technology, Jalandhar, India
[b] Department of Electronics and Communication Engineering, Guru Nanak Dev University, Regional Campus, Jalandhar, India

ARTICLE INFO

*Article history:*
Received 31 March 2015
Revised 13 October 2016
Accepted 13 October 2016

*Keywords:*
Electrocardiogram
ECG
Pre-processing
Data compression
Sample entropy
SampEn
JPEG2000
Quality score
Complexity sorting

ABSTRACT

This paper proposes an effectual sample entropy (SampEn) based complexity sorting pre-processing technique for two dimensional electrocardiogram (ECG) data compression. The novelty of the approach lies in its ability to compress the quasi-periodic ECG signal by exploiting the intra and inter-beat correlations. The proposed method comprises the following steps: (1) QRS detection, (2) Length normalization, (3) Dc equalization, (4) SampEn based nonlinear complexity sorting and (5) Compression using JPEG2000 Codec. The performance has been evaluated over 48 records from the MIT-BIH arrhythmia database. The average quality score (QS) measurements at different residual errors were 42.25, 4.73, and 2.75 for percentage root mean square difference (PRD), PRD1024, and PRD Normalized respectively. The work also reports extensive experimentations on the compressor for various durations of the ECG records (5–30 min, with 5-min increment). The proposed algorithm demonstrates significantly better performance in comparison to the contemporary state-of-the-art works present in the literature.



## 1. Introduction

The incorporation of wireless communication technology in the field of the tele-cardiological platform has played a vital role in the timely monitoring of electrocardiogram (ECG) signals, especially in remote areas. The ECG signal that monitors the electrical activity of the heart is normally characterized by its different clinical features based on set points (P, QRS, T) and intervals (PR-interval, QT-interval, and RR-interval) that reflect the rhythmic electrical repolarization and depolarization of the ventricles and atria [1].

The redundant information in an ECG signal is due to the presence of inter, and intra-beat correlation. Broadly, ECG data compression methods fall under one of three classifications namely, direct, transformed, or parameter extraction method [1]. The direct method reduces ECG sample points in the time domain. This category includes turning point (TP), amplitude zone time epoch coding (AZTEC), improved modified AZTEC [2], coordinate reduction time encoding, the Fan algorithm, and the ASCII character based encoding systems [3,4]. The transformed method analyses energy distribution or compaction by persuading in other domains, which include the Fourier transform, the Fourier descriptor, the Karhunen–Loeve transform (KLT) [1], the Walsh transform, the discrete cosine transform (DCT), DCT with modified stages [5] and the wavelet transform

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. M. R. Daliri.

* Corresponding author.
*E-mail address:* anukul66@gmail.com (A. Pandey).

http://dx.doi.org/10.1016/j.compeleceng.2016.10.012
0045-7906/

[6]. The compressed sensing model [7], the fractal model [8] and the accuracy driven sparse model [9] were recently introduced. The parameter extraction method is based on the dominant extracted features from the raw ECG signal. These are neural or syntactic, peak picking and linear prediction [10] based methods.

Numerous researchers have reported ECG data compression procedures by formulating two-dimensional (2D) arrays from ECG signals to better exploit the inter- and intra-beat correlations by the employed encoder. [11–13]. The "cut and align beats approach with 2D DCT" along with "period normalization and truncated SVD algorithm" are the existing pre-processing techniques that achieve decent ECG compression results. The state-of-the-art image encoders, such as JPEG2000, usually employ this type of a pre-processing step. The pre-processing method proposed here is a modification of the technique presented by [11]. In [12], the authors proposed a lossy ECG compression method based on translating the ECG signal into a 2D ECG array. After translation, a period sorting pre-processing approach was applied, which is composed of a length-based organization of all the periods. The inter-beat correlation of the 2D ECG has been exploited by Chou et al. [12]. These researchers formulated a 2D ECG array and sorted in ascending order according to period length. Then, inter-beat correlation among similar period lengths was perceived. In the analysis by Filho et al. [11] it is verified that this assumption does not hold for the large set of ECG signals of pathology subjects. One more pre-processing method consists of a QRS detector, period length normalization, Dc equalization, complexity sorting and image transformation as designed in [11]. The technique centred on the pre-processing stage by reducing the high-frequency content in the vertical direction of the equalized 2D ECG array.

Although the vertical, high-frequency content of the formulated 2D ECG array has been reduced by Dc equalization and variance based linear complexity sorting techniques, adjacent lines may be quite dissimilar by following this pre-processing mechanism. In some circumstances, it fails to assure the maximum similarity through adjacent periods. As an ECG is a quasi-stationary signal, the variance based linear complexity sorting operation used in [11] can be further improved by nonlinear processing typically using entropy based complexity sorting. Physiological systems, such as the cardiovascular system, are composed of many nonlinearly interacted interdependent subsystems, resulting in highly complex signals. Many methods have been used to evaluate complexity, using both linear and nonlinear measures. Methods based on linear modelling are widely known and standardized. However, as a complex system, the cardiovascular system might be better assessed by nonlinear methods. Sample entropy (SampEn) represents the nonlinear quantitative metric used to quantify the irregularity of a time series [14]. The main objective of the present paper is to explore the SampEn based nonlinear complexity indices for sorting and rearrangement of the rows of the 2D ECG matrix array from the simplest to the most complex ones.

The rest of this paper is organized as follows: Section 2 consists of the database and performance measure description; justification to use the sample entropy based sorting is discussed in Section 3. Section 4 describes the proposed 2D ECG processing mechanism; Section 5 demonstrates the experimental results; finally, Section 6 presents the paper's conclusions.

## 2. ECG database and performance measures

The Massachusetts Institute of Technology (MIT) provides useful resources for various research works, such as the analysis, processing, and compression of different signals. These resources incorporate with various databases, containing the physiological data and software for creating, viewing and analysing such recordings. In the present study, ECG samples from the Arrhythmia Database [15] (http://ecg.mit.edu) were used to evaluate the proposed method. The database contains 48 approximately thirty minutes two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory. The 11-bit resolutions over 10 mV ranges with sampling frequency of 360 Hz per sample channel were recorded. Tests were performed with the second channel of all 48 records of this database.

To access the proposed pre-processing technique's efficiency, 21,600 samples (10 min duration) were used and evaluated with the following data compression metrics [5,9,10,16]. Performance statistics were also computed for varying durations of the ECG records (5, 10, 15, 20, 25 and 30 min) thereby indicating the suitability of the designed approach for short, medium and long term ECG data.

The various performance metrics are defined below.

### 2.1. Compression ratio (CR)

The CR provides information of the extent by which the compressor eliminates the redundant data. The higher the CR, the fewer number of bits needed to store or transmit the data. CR is defined as follows:

$$CR = \frac{B_0}{B_c} \tag{1}$$

where, $B_0$ is the total number of bits required to represent the original data and $B_c$ is the total number of bits required to represent the compressed data along with the side information needed for retrieval of the original data. CR can also be defined as follows:

$$CR = \frac{11 \times f_s \times N}{B_{c2d} + B_s} \tag{2}$$

where $f_s$ is the sampling frequency, $N$ is the total number of samples in a one dimensional (1D) ECG signal, $B_{c2d}$ is the total number of bits in the compressed (2D) ECG and $B_s$ is the bits required to store the side information.