Measurement 45 (2012) 1651–1660

Contents lists available at SciVerse ScienceDirect

Measurement

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

An ECG signal compression technique using ASCII character encoding

S.K. Mukhopadhyay [a], S. Mitra [b], M. Mitra [a,*]

[a] Department of Applied Physics, Faculty of Technology, University of Calcutta, 92, APC Road, Kolkata 700 009, India
[b] Department of Electronics, Netaji Nagar Day College (Affiliated to University of Calcutta), 170/436, N.S.C. Bose Road, Regent Estate, Kolkata 700 092, India

ARTICLE INFO

Article history:
Received 2 June 2010
Received in revised form 4 November 2011
Accepted 23 January 2012
Available online 10 February 2012

Keywords:
Sign bit generation
Searching the largest number and amplification
Grouping
Ungrouping
De-amplification

ABSTRACT

Efficient software based new ECG data compression algorithm is developed here. The whole algorithm is written in the C-platform. The algorithm has been applied to various ECG data of all the 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6), taken from PTB Diagnostic ECG Database (ptbdb). In this method, only the "Voltage" values are gathered from the corresponding input ECG data and then those are multiplied by a considerably large integer number to convert them into integer. In the next step, these integer numbers are grouped maintaining some logical criteria. This grouping is functioning as the main compression technique. In this technique the grouping has been done in such a way that every grouped number resides under valid ASCII value. Then all the grouped numbers along with sign bit and other necessary information are converted into their corresponding ASCII characters. It is observed that this proposed algorithm can reduce the file size significantly. The data reconstruction algorithm has also been developed using the reversed logic and it is seen that data is reconstructed preserving the significant ECG signal morphology. Basically, ECG is a voltage–time characteristic and since the sampling time is constant, the "time" data can be automatically reconstructed at the receiving end. Therefore concentration has been given only to compress the 'Voltage' values.

1. Introduction

Electrocardiogram (ECG) is a non-invasive, transthoracic diagnostic technique, which gives an inkling of the electrical activity of human heart over time. The shape and size of the P-QRS-T wave and the time intervals between various peaks contains useful information about the nature of disease afflicting the heart. Since bio-signals are highly subjective, the symptoms may appear at random in the timescale. Hence monitoring of ECG data may be extended over a long period of time resulting in a huge amount of data for storage or transmission. Hence reduction of data storage place without much affecting the clinical significance of the ECG signals is mandatory for most of the compression methods reported in the literature. The ECG data compression is also required for the transmission of ECG signals across public telephone networks, cellular networks and wireless communication systems ("telemetry"). But the overall goal for all the compression techniques is to represent a signal as accurately as possible using the fewest number of bits, by applying either lossless compression, in which the reconstructed signal is an exact replica of the original signal, or lossy compression, where the reconstructed data may well be different from the original, but is "close enough" to be useful in some way. Lossy data compression is used frequently on the Internet and especially in streaming media and telephony applications. The advantage of lossy methods over lossless methods is that in some cases a lossy method can produce a much smaller compressed file than any known lossless method, and still meeting the requirements of the application. With lossy compression, a certain amount of distortion has to be accepted in the reconstructed signal, although the distortion must remain small

* Corresponding author.
E-mail addresses: sonartoritag@yahoo.co.in (S.K. Mukhopadhyay), susa68@hotmail.com (S. Mitra), madhuchhanda94@rediffmail.com (M. Mitra).

0263-2241/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.measurement.2012.01.017

enough not to modify or jeopardize the diagnostic content of the ECG. In addition to the previous categorization, ECG compression algorithms can also be broadly classified into (i) direct methods, (ii) transformed, and (iii) parameter extracted. Performance summaries of the ECG compression methods can be found in some literatures including [1,2]. Direct methods like AZTEC [3], DPCM [4], Turning Point (TP) [5], CORTES [6], SAPA/FAN [7], etc., generally retain samples that contain important information about the signal discarding the rest. The basic idea of all these algorithms is to break the ECG signal into consecutive linear segments. This is done by selecting a set of significant samples in the encoder. Among transform schemes, the wavelet transform [8–12] is the most effective one because of its good localization properties in the time and frequency domain. Unlike the STFT, the time–frequency kernel for the WT-based method can better localize the signal components in time–frequency space. Beside wavelet, Fourier [13], Walsh [14], KLT [15] and discrete cosine transform (DCT) are also used as transform schemes. On the other hand parameter extraction methods are mainly based on linear prediction and long-term prediction methods [16]. It is an irreversible process, where a preprocessor is employed to extract some features that are later used to reconstruct the signal. In recent years, few mathematical model based methods like Jacobi Polynomials [17], Discrete Sine Interpolation (DSI) [18], Cut N Align Beats (CABs) approach [19], etc. are also reported. In CAB based approach, a two dimensional (2-D) transform is performed. This 2-D transform method utilizes the fact that ECG signals generally show two types of redundancies—between adjacent heartbeats and between adjacent samples. A heartbeat data sequence is cut and beat-aligned to form a 2-D data array, which are then processed by 2D-DCT data compression method. Whereas in DSI method, DFT is used efficiently to achieve a Discrete Sine Interpolation (DSI) function for getting both compressed and decompressed ECG data. Many authors have tackled the problem of ECG compression by treating it as an image [20–23], allowing the encoder to efficiently exploit the inter and intrabeat dependencies. Another efficient ECG signal compression based on local extreme extraction, adaptive hysteretic filtering and Lempel–Ziv–Welch (LZW) coding [24] is also reported.

In the present work we have introduced an effective protocol for ECG data compression. A group of 16 data (eight time values and eight voltage values) is taken at a time from the input ECG data file. Then only the voltage values are sorted for later processing. Starting from first voltage value if anyone is found to be negative then one (1) and for positive, a zero (0) is taken as the symbol of sign for the corresponding voltage values. The decimal equivalent to these eight 0s and 1s is evaluated as the sign bit of the corresponding eight voltages. These voltages are then multiplied by a suitable large number to obtain integer number. Those integers are now grouped maintaining some essential logical criteria except few critical numbers that are reserved as some special characters and are treated differently. At last these grouped integers along with some essential information (sign bit, amplification factor, etc.) are printed in their ASCII characters.

In the reconstruction module of the algorithm, this data is decompressed using reverse programming logic by separating the sign-bit, amplification factor and other information. The output file of the reconstructed ECG signal is also generated as a time–voltage series.

2. ECG data compression and reconstruction protocol

Proposed algorithm is divided into following two sections: (i) data compression (ii) data reconstruction. The data compression algorithm is referred to as *'Electrocardiogram Data Compression by Character Encoding (EDCCE)'* and the data reconstruction algorithm is referred to as the *'Electrocardiogram Data Reconstruction by Character Decoding (EDRCD)'* in the rest of the literature.

2.1. Programming logic of EDCCE

Block schematic of the proposed "EDCCE" algorithm is given in Fig. 1.

2.1.1. Reading ECG data from input file

The first objective is to read only the 'Voltage' values from the 'time–voltage' input ECG data file we intend to compress. All the ECG data files are taken from PTB diagno-

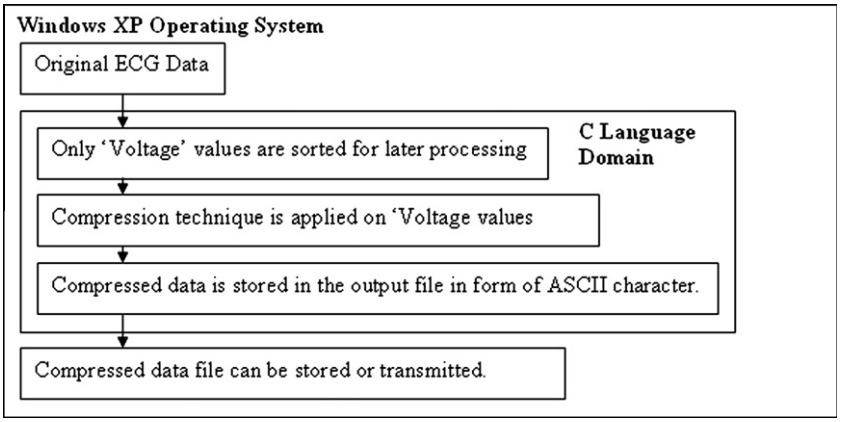

Fig. 1. Block schematic of proposed "EDCCE" algorithm.