2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)
21 - 23 Dec 2017, Dhaka, Bangladesh

ECG signal Compression using Data Extraction and Truncated Singular Value Decomposition

Syed Salman Kabir*, Mamshad Nayeem Rizve, Md. Kamrul Hasan
Foysal Department of Electrical & Electronic Engineering
Bangladesh University of Engineering & Technology, Dhaka, Bangladesh
*salmankabir.buet@gmail.com

***Abstract*—In this paper a novel algorithm is used for compressing ECG data. The algorithm uses two techniques simultaneously to get a better compression of ECG data. Firstly Data extraction: a sampling technique based on a user defined threshold value is applied on ECG followed by Truncated Singular Value Decomposition. Changing the number of Singular Values higher compression can be achieved. The proposed method provides two variables: Threshold & Number of Singular Value, which can be changed to achieve optimum compression while keeping the error low. Compression performance & quality of reconstructed ECG in the proposed method is evaluated using Compression Ratio (CR) & Percent Root Mean Square Difference (PRD).**

***Index Terms*—Data Extraction, Truncated Singular Value Decomposition (TSVD)**

I. Introduction

Electrocardiogram (ECG) signal is used to diagnose heart disease in a patient. In an ambulatory monitoring system, the volume of data is necessarily large, as a long period of time is required in order to gather enough information about the patient. Therefore, an effective data compression scheme for ECG signals is required in many practical applications including: ECG data storage, Ambulatory recording systems, transmission of ECG data over telecommunication network. ECG signals are recorded from patients for both monitoring and diagnostic purposes. Therefore, the storage of computerized ECG data becomes necessary. The effective use of wired or wireless communications resources requires a real-time data compression and transmission method in the case of a real-time ECG monitor or multichannel bio-signal acquiring devices. But ECG signal is usually bulky, so how to make it's stored and transmission quantity smaller to improve real-time process, has become a widely researched field. The goal of ECG compression techniques is to achieve a reduced information size, while preserving the relevant diagnostic information in the reconstructed signal. A data compression algorithm should allow reconstruction of the data with acceptable fidelity. The main objective of any compression technique is to remove redundancy and achieve maximum data volume reduction while preserving the necessary diagnosis features. Data compression methods have been extensively discussed and classified into following major categories[1]: Parameter extraction techniques (average beat subtraction[2], cycle pool based compression), Direct data compression techniques (DPCM, AZTEC, SAPA), Transformation methods (KL Transform [3], Wavelet Transform[4]). Again compression techniques can be categorized as lossless & lossy compression. The method proposed here utilizes the redundancy of ECG signal and its periodicity. While keeping the significant portion of ECG signal (QRS complex) unaffected, it tries to reduce data from redundant parts. This idea is identified as Data Extraction Method. After that the ECG signal is divided into R-R segments to form a matrix on which Singular Value Decomposition Technique is applied. In order to get higher compression lesser number of singular values is stored. To reconstruct the signal a position matrix is used which helps separate the data omitted and the data initially saved. The performance of the proposed algorithm, in terms of reconstructed signal quality and compression ratio, is evaluated using the MIT -BIH Arrhythmia Database.

II. Proposed Method

A. Data Extraction

In this method the modulus of amplitude difference between two consecutive points is measured and compared with a threshold value. If the amplitude difference is smaller than the threshold value, the first data point is omitted. If the amplitude difference is equal or bigger than the threshold value, the first data point is stored. As a result less number of data is saved in flat segments of the signal (PR segment, ST segment) while the QRS section is kept unaffected.

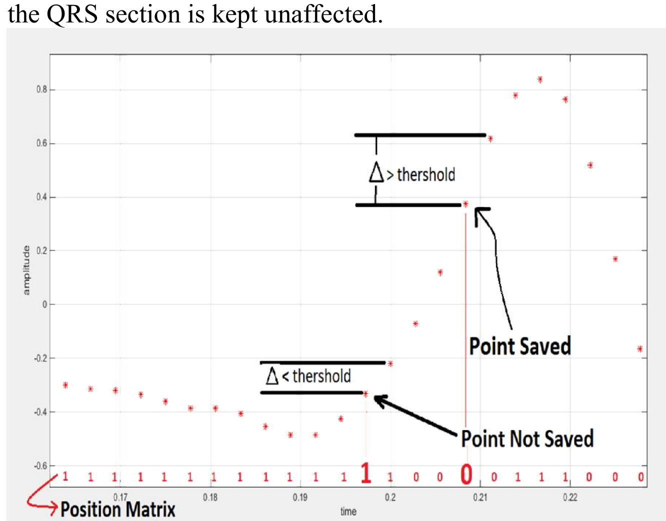

Figure 1 Overview of Data Extraction method

Thus the redundancy of data can be minimized. In this method some points are saved and some are omitted so it is necessary to keep track of the position of the points which are being

978-1-5386-2175-2/17/$31.00 ©2017 IEEE

saved because the position is needed to reconstruct the signal using cubic-spline interpolation [5]. To do so '0' was marked as the position of the points being saved while marking the rest as '1' which gives us an array of 1's and 0's, called position matrix. For better compression this position matrix is created instead of saving the actual position.

B. Truncated Singular Value Decomposition

The method of singular value decomposition (SVD) has been applied in a wide range of areas, such as image compression, texture processing, and feature extraction. A prominent feature of SVD is the separation of the fundamental structural modes constituting a system that can be used to analyze quasi-periodic processes. After reducing the number of data points using Data Extraction method, SVD is applied on the ECG data. Consider an ECG signal (t) , which has p consecutive periods each having length q. These periods were formed in a way that it starts from one R-wave and ends at the next one[6]. $m(t)$ can then be rearranged as a two-dimensional matrix M, which can be expressed as

$$M = \{m_i(t) | i = 1,2 \ldots..p, t = 1,2 \ldots \ldots q\}$$

$$= \begin{bmatrix} m(1) & m(2) & \cdots & m(q) \\ m(q+1) & m(q+2) & \cdots & m(2q) \\ \vdots & & \ddots & \vdots \\ m((p-1)q+1) & m((p-1)q+2) & \cdots & m(pq) \end{bmatrix}$$

Every row of matrix M represents one period of ECG signal. Where $m_i(t)$ is the i-th period of $m(t)$. Therefore, the SVD of the p× q matrix M can be performed as $M = UDV^T$, where $U \in R^{p\times p}$ $V \in R^{q\times q}$ are the left and right singular vectors, respectively. The $p \times q$ matrix $D = [diag\{x_1 x_2 \ldots \ldots x_N\}: 0]$ and $x_1 x_2 \ldots \ldots x_N$ are the singular values of matrix M where $x_1 > x_2 > \cdots > x_N$ and $N = \min(p, q)$. According to the decomposition processes of SVD, the matrix M, which is filled with the repetitive pattern of consecutive rows, can be decomposed into a set of vector v_i and associated with vectors $u_i x_i$, where i=1,2…..R . All the singular values will be zero, except x_1 with v_1 representing the principal component and $u_1 x_1$ is the scaling element representing the amplitude over the entire period if the matrix M is a rank-one matrix which means $m(t)$ becomes an exact periodic signal with length q. So only one singular value x_1 and related singular vector pair $\{u_1 v_1\}$ is sufficient to recover the original matrix using the following expression

$$\widehat{M} = u_1 x_1 v_1^T$$

Original waveform can be recovered from $\widehat{M}$ after proper matrix conversion. As ECG is a practical periodic signal all the periods will not contain same values like a sinusoid and there will be differences to a certain extent. ECG is a combination of periodic or nearly periodic signals and some aperiodic signals. Consequently, M will become a full rank matrix. But the main information can still be recovered from the first few Singular values if

$$\frac{x_1^2}{x_i^2} \gg 1 \; where \; i = 2,3,4 \ldots \ldots N$$

If r of N singular values are much greater than the rest, then truncating the singular value matrix D, M can be recovered using

$$\widehat{M} = U'D'V'^T$$

where $U' \in R^{p\times r}$ with orthonormal columns, $D' = R^{r\times r}$ diagonal and $V' \in R^{q\times r}$ with orthonormal columns. The truncated U', D', V' matrices contain less data than their original counterparts. These matrices with the position matrix from data extraction method are saved as compressed data and used to reconstruct the ECG signal.

C. Signal Reconstruction

After estimation of the ECG data using aforementioned truncated SVD, the matrix $\widehat{M}$ is rearranged in an array from its R-R segment. Previously redundant data were removed using Data Extraction technique and the position of the removed data was kept in the position matrix. With the help of position matrix the signal is reconstructed by cubic spline interpolation. Figure 3 shows the original and reconstructed signal of sample number 100.m of MIT -BIH Arrhythmia Database. No significant anomaly is visible in the reconstructed signal.

Figure 2 Flow chart of Proposed Algorithm

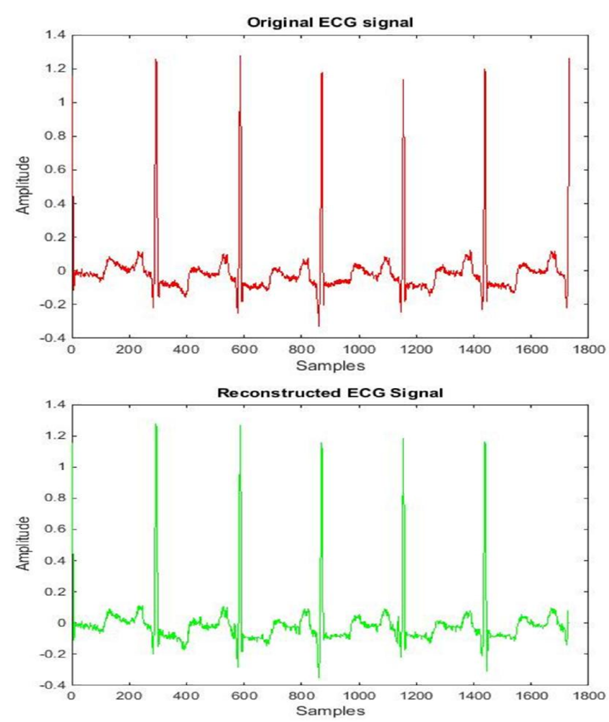

Figure 3 Original & Reconstructed Signal using proposed method

III. Result Analysis

The performance of the proposed method is analyzed based on two parameters: Compression Ratio (CR) & Percent Root-

Mean-Square Difference (PRD). CR is the ratio of the original data to compressed data without taking into account any factors such as bandwidth, sampling frequency, data size and selection of lead. It is given by:

$$CR = \frac{original\ file\ size}{compressed\ file\ size}$$

Percent Root-Mean-Square Difference (PRD) is a measure of error loss. This measure evaluates the distortion between the original and the reconstructed signal. PRD calculation is as follows:

$$PRD = \sqrt{\frac{\sum_{i=1}^{n}(ORG(i) - REC(i))^2}{\sum_{i=1}^{n}(ORG(i))^2}} \times 100\%$$

For different threshold values different number of singular value was taken and the resultant CR & PRD are given for ECG sample 100m form MIT -BIH Arrhythmia Database in Table 1.

Table 1 CR & PRD For Different No. Of Singular Values

No of Singular Values	Threshold=0.01		Threshold=0.005	
	CR	PRD(%)	CR	PRD(%)
6	*6.27*	*3.54*	*5.88*	*2.02*
5	*6.63*	*5.03*	*6.24*	*3.68*
4	*8.01*	*9.08*	*7.88*	*6.71*
3	*9.5*	*13.92*	*9.3*	*11.57*
2	*11.12*	*18.9*	*10.59*	*14.9*

The proposed method uses two different variables to change the CR & PRD values and a better CR can be achieved while keeping the PRD minimum. In [6] CR value is calculated changing bit rate where bit rate is kept constant throughout the whole process in this paper. For different samples of MIT-BIH Arrhythmia Database the proposed algorithm was tested and the results show significant correlation. The CR values for different samples are illustrated in figure-4. The result from

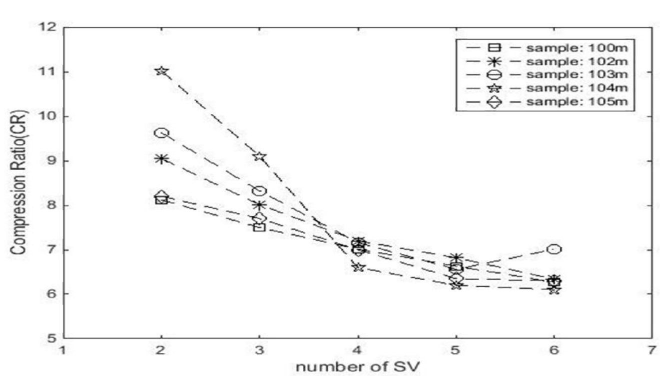

Figure 4 CR vs Number of SV

the simulation is then compared with other existing methods of ECG compression based on CR & PRD. Table-2 shows that the proposed method improves the CR significantly & further improvement can be achieved by increasing threshold value.

Table 2

Method	CR	PRD%
CORTES [7]	4.8	7
DPCM with entropy coding[8]	7.8	3.5
Fourier Descriptors[9]	7.4	7
Proposed method	11.12	18.9

IV. CONCLUSION

The paper proposes an algorithm with integration of two techniques to get high compression ratio, low PRD and two completely different variables (Threshold value & Number of Singular Value) to get the desired result. The reconstruction process includes using the position matrix and cubic spline interpolation which is simple and easy to perform. Using the algorithm on different samples of MIT-BIH database, it can be inferred that the algorithm gives similar performances despite the nature of the ECG signal.

References

[1] S. M. S. Jalaeddine, C. G. Hutchens, R. D. Strattan, and W. A. Coberly, “ECG compression techniques, A unified approach,” IEEE Trans. Biomed. Eng., vol. 37, pp. 329–343, Apr. 1990.

[2] S. Pathoumvanh, S. Airphaiboon and K. Hamamoto, "ECG Data Compression using Adaptive Beat Subtraction Method," *2008 International Symposium on Communications and Information Technologies*, Lao, 2008, pp. 477-481.

[3] S. Olmos, M. Millan, J. Garcia and P. Laguna, "ECG data compression with the Karhunen-Loeve transform," *Computers in Cardiology 1996*, Indianapolis, IN, USA, 1996, pp. 253-256.

[4] B. A. Rajoub, “An Efficient Coding Algorithm for the Compression of ECG Signals Using the Wavelet Transform,” IEEE Transactions on Biomedical Engineering, vol. 49, no. 4, pp. 355–362, Apr 2002.

[5] S. A. Dyer and J. S. Dyer, "Cubic-spline interpolation. 1," in *IEEE Instrumentation & Measurement Magazine*, vol. 4, no. 1, pp. 44-46, Mar 2001.

[6] Jyh-Jong Wei, Chuang-Jan Chang, Nai-Kuan Chou and Gwo-Jen Jan, "ECG data compression using truncated singular value decomposition," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 5, no. 4, pp. 290-299, Dec. 2001.

[7] J. P. Abenstein, W. J. Tompkins, "New data-reduction algorithm for real-time ECG analysis", *IEEE Trans. Biomed. Eng.*, vol. BME-29, pp. 43-48, Jan. 1982.

[8] U. E. Ruttimann, H. V. Pipberger, "Compression of the ECG by prediction or interpolation and entropy encoding", *IEEE Trans Biomed. Eng.*, vol. BME-26, pp. 613-623, Nov. 1979.

[9] B. R. S. Reddy, I. S. N. Murthy, "ECG data compression using Fourier descriptors", *IEEE Trans. Biomed. Eng.*, vol. BME-33, pp. 428-434, Apr. 1986.