New Improved Methodology for ECG Signal Compression

Rupali V Tornekar[1], Suhas S Gajre[2]

[1]Vishwakarma Institute of Technology, Pune, India
[2]Shri Guru Gobind Singhaji Institute of Engineering & Technology, Nanded, India

Abstract

ECG of a person is being recorded for diagnosis of heart diseases, regular checkups, fitness and many other diseases also. Hence, huge amount of ECG data is being generated daily in hospitals. ECG recording and monitoring consume lots of memory space of digital computers. Data compression plays a vital role in reducing storage space and utilizing transmission bandwidth effectively. Objective of the research work is to propose a robust and effective method for ECG signal compression.

This paper includes extraction of key morphological features, statistical features from decomposed signal, analysis in wavelet domain and classification of feature set. To reduce dimensionality of feature set, principle component analysis (PCA) is applied. Accuracy achieved with 15 principle components is same as pure wavelet transform with significant improvement in compression ratio (CR) by a factor of 4:1. MITBIH Arrhythmia database is used for experimentation. Accuracy obtained with K nearest neighborhood (KNN) classifier is 82.60% and sensitivity 92.30%. Research work carried out here is an improvement in the existing telemedicine technology.

1. Introduction

As per recent study by the Registrar General of India (RGI) and the Indian Council of Medical Research (ICMR), nearly about 25 percent of deaths occur because of heart diseases. Now a day, telemedicine has taken shape in biomedical signal processing field. Since patients are able to get quick decisions from medical experts, it helps to save life of people in emergency, remote places. Electrocardiograph (ECG) is a physiological parameter which represents electrical activity of Heart. A 12 channel ECG recording and monitoring system requires nearly 512MB of storage disks daily for signals sampled at 360Hz sampling frequency and 11 bit resolution. As digital technology is moving ahead at a rapid pace, data storage is becoming a critical issue. ECG compression is active area of research in biomedical field.

It has been observed that many algorithms are presented in the literature for compression of ECG signals. They are categorized into three different types: Direct time domain, transform based and parametric techniques [1][2][3]. In recent years wavelet based compression methods have taken sharpen edge especially in biomedical signal processing applications.

The optimal zonal wavelet-based ECG data compression (OZWC) algorithm [1] achieved a maximum compression ratio of 18: 1 with low percentage root mean square difference (PRD). The wavelet filters based on beta function [4] and their derivatives, improve the compression of signal as well as minimize distortion. the optimal zonal wavelet coding (OZWC) method and the wavelet transform higher order statistics-based coding (WHOSC) can contribute to and enhance the medical data compression performance suitable for a hybrid mobile telemedical system especially in low bandwidth mobile systems [5]. Multiscale principal component analysis (MSPCA) based on average fractional energy contribution of eigenvalue in a data matrix [6] is implemented using uniform quantizer and entropy coding of PCA coefficients. compression ratio of 5.98:1 is found with PRD value 2.09% .

Raeiatibanadkooki, et al. proposed a compression method by focusing on compressing the ECG signal with no loss of essential data and also encrypting the signal to keep it confidential from everyone, except for physicians. Mobile processors were used and there was no need for any computers to serve this purpose [7]. After initial preprocessing such as removal of the baseline noise, Gaussian noise, peak detection and determination of heart rate, the ECG signal is compressed. Then, Huffman coding with chaos for compression and encryption of the ECG signal were used.

It is found that there is a scope to study the effect of CR and PRD on obtained features and also sensitivity, accuracy after classification to make a closed loop system for validation of compression algorithms. Proposed work focuses on Compression of ECG signal, extraction of features from decompressed signal, principle component analysis for further data compression and validation using a classifier.

In telemedicine, recorded data of patients is

transferred over long distances for medical expert's opinion. Proposed work is an improvement in the existing telemedicine technology as it also provides first aid analysis to both medical experts and patients. It will help experts in taking quick decisions. It may save life of human being in emergency cases, heavy rain, traffic, rural areas, long distance healthcare unavailability. Compression is required for reducing memory size, cost and time of data transmission. Availability of bandwidth is crucial in rural areas. Main aim of compression is to utilize bandwidth effectively. This research paper focuses on comparative study of lossy Compression methods based on wavelet and PCA. For validation of heart abnormality, accuracy, sensitivity and specificity parameters are calculated. For classification KNN classifier is used.

2.   Work Methodology

This paper demonstrates the new improved ECG signal lossy compression methodology based on wavelet and principle component analysis (PCA).

2.1.   Preprocessing

ECG signal is preprocessed first before applying compression method. Preprocessing consists of normalization, mean removal. Due to normalization and mean removal, the number of significant wavelet coefficients gets reduced and maximum amplitude is made less than one.

2.2.   Compression algorithm

ECG signal is compressed using wavelet transform [8][9]. One of the Best Choice of mother wavelet [10][11] for ECG signal processing is Bior4.4. This mother wavelet is chosen on the basis of filter length, matching of shape of mother wavelet with the signal to be decomposed and energy spectra range. QRS complex normal frequency range of 20 Hz to 60 Hz lies in the fourth decomposition level. So signal is decomposed into approximation and detail coefficients upto fourth level.

2.3.   Feature extraction

Key Morphological features include Q, R and S detection, amplitudes and intervals [12], heart rate. After Reconstruction of the ECG signal, fourth level detail coefficients are extracted. To find R peak positions, zero crossing detection of first order derivative of these fourth level detail coefficients is performed. The correct location of R peaks on original time domain signal is found by multiplying the extracted R peak positions with a scale factor 2^4. The exact R peak is located by finding maximum peak in chosen sliding window. Once the correct R peak positions are located, then Q and S peak positions are found as the highest negative peaks in a sliding window.

Mean, mode, variance and higher order statistical measures like skewness and kurtosis of approximation and detail coefficients are also calculated. Standard Deviation gives the measure to quantify the amount of variation or dispersion of a set of data values from the mean. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. Data with high kurtosis tend to have a distinct peak near the mean, decline rapidly and has heavy tails. Skewness indicates asymmetry and deviation from a normal distribution.

2.4.   Classification

KNN classifier is used for to check heart abnormality. K-Nearest Neighbor (KNN) classification is a very simple, yet powerful classification method. It became popular due to its simplicity and relatively high convergence speed. When dealing with continuous attributes the difference is calculated as the Euclidean distance. If the first instance is $(a_1, a_2, a_3 \ldots a_n)$ and the second instance is $(b_1, b_2, b_3 \ldots b_n)$ the distance between them is calculated by the following formula:

$$D = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \ldots (a_n - b_n)^2} \quad (1)$$

Continuous attributes are normalized to have the same influence on the distance measured between instances. The weighing scheme of the class numbers is often a majority rule, but other schemes are conceivable. The number of the nearest neighbor's k, should be odd in order to avoid ties, and it should be kept small, since a large k tends to create misclassifications unless the individual classes are well-separated. The performance of a KNN classifier is always at least half of the best possible classifier for a given problem. One of the major drawbacks of KNN classifiers is that the classifier needs all available data. This may lead to considerable overhead, if the training data set is large. 50% dataset is used for training.

2.5.   Performance measures

Quantitative measures of performance used are accuracy, sensitivity and specificity. The detection is said to be true positive (TP) if the algorithm correctly detects the normal signal, true negative (TN) if the algorithm correctly detects the abnormal signal, false negative (FN) if algorithm detects normal signal as abnormal, and false positive (FP) if the algorithm detects abnormal signal as normal. The parameters which are used to evaluate the

performance of classifier are: sensitivity, specificity and accuracy [9][14]. Average Performance parameters are calculated after applying on all 47 records of MITBIH database.

They are calculated using following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (3)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (4)$$

$$PRD = \sqrt{\frac{\sum_{n=1}^{N}(x(n) - x'(n))^2}{\sum_{n=1}^{N} x^2(n)}} \qquad (5)$$

x(n)- original signal

x'(n)- reconstructed signal,

N- Length of the signal over which the PRD is calculated.
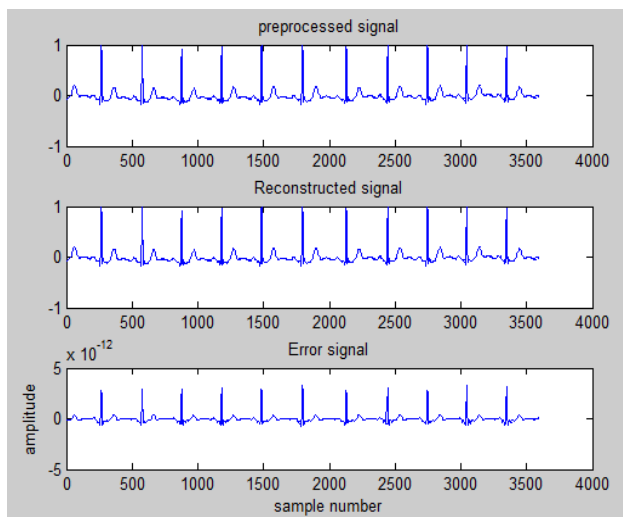
3.　　Experimental Results

Figure1: Error signal after compression of ECG using wavelet transform for record 103 from MITBIH database

The compression algorithm is tested on all records of MITBIH Arrhythmia database. Lead II Recordings of 10 sec of 47 patients with sampling frequency 360 Hz and 11 bit resolution is used for experimentation. ECG signal decomposed upto 4th level using Bior4.4 mother wavelet.

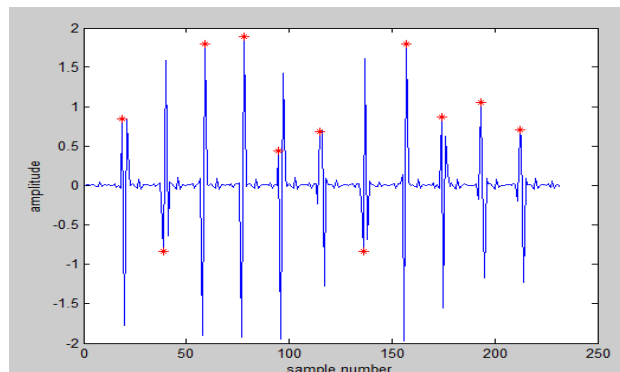Reconstruction error is very less in the order of 10^{-12} as shown in Figure1.

Figure2: zero crossing detection of first order derivative of fourth level detail coefficients

R peak positions are found and marked on 4th level detail wavelet coefficients as shown in figure2. After multiplying these positions by 2^4 and using a suitable sliding window, Exact Q, R and S peak positions are found and marked on reconstructed ECG signal as shown in Figure3.
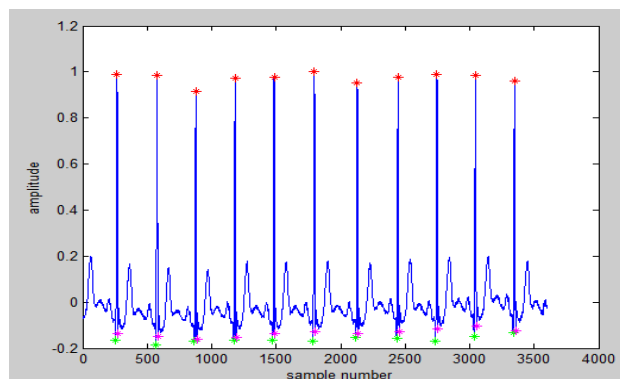

Figure3: QRS detection on Record number 103 of MITBIH Arrhythmia database

Research work presented in this paper is compared with existing algorithms mentioned in [13] in terms of CR and PRD. Table 1 clearly shows PRD value in case of proposed algorithms is much less than others.

Table1. CR and PRD comparison of different existing algorithms

Algorithm	Record	CR	PRD(%)
Lu et. al	117	8:1	1.18
Hilton	117	8:1	2.6
Djohan	117	8:1	3.9

et.al			
Wei et.al	117	10:1	1.18
Proposed	**117**	**8:1**	**2.2389e-10**
ASEC	119	21.6:1	5.5
Lu et.al	119	21.6:1	5.0
Proposed	**119**	**8:1**	**2.3126e-10**

To reduce dimensionality of feature set, principle component analysis (PCA) is applied. Performance of wavelet transform method is compared with PCA of wavelet coefficients having 15 and 5 components as best and worst case respectively. By considering 5 principle components, classification accuracy is degraded by 2 % as that of wavelet transform. Sensitivity is very important evaluation parameter in biomedical signal processing and medical diagnosis. Accuracy obtained with KNN classifier is 82.60% and sensitivity 92.30%. Proposed work is an improvement in the existing telemedicine technology.

Table2. Performance comparison

parameter	Wavelet	Wavelet+ PCA(15)	Wavelet+ PCA(5)
Compression method	Lossy	**Lossy**	Lossy
Compression ratio	4:1	**8:1**	14:1
Computational complexity	N^2	$N^2+NM^2+N^3$	$N^2+NM^2+N^3$
Accuracy%	82.60	**82.60**	80.43
Sensitivity%	92.30	**92.30**	88.88
specificity%	70	**70**	68.42

4. Conclusion

In this Research work, key morphological features of ECG signal and statistical features of wavelet coefficients are considered for classification. Principle Component Analysis implementation on wavelet features gives satisfactory results in terms of compression ratio, accuracy of classification and sensitivity. By considering 15 principle components, classification metrics give same results as wavelet with significant improvement in compression ratio by a factor of 4:1 as shown in Table2. Proposed Work is an improvement in the telemedicine as it includes compression of data and first aid analysis.

5. References

[1] Istepanian, Robert S. H., and Arthur A. Petrosian. "Optimal zonal wavelet-based ECG data compression for a mobile telecardiology system." Information Technology in Biomedicine, IEEE Transactions on 4.3 (2000): 200-211.

[2] Khanam, Ruqaiya, and Syed Naseem Ahmad. "ECG Signal Compression for Diverse Transforms." Information and Knowledge Management. Vol. 2. No. 5. 2012, pp. 1-9

[3] Blanchett, T., G. C. Kember, and G. A. Fenton. "KLT-based quality controlled compression of single-lead ECG." IEEE Transactions on Biomedical Engineering 45.7 (1998): 942-945.

[4] Kumar, Ranjeet, A. Kumar and Rajesh K. Pandey. "Beta wavelet based ECG signal compression using lossless encoding with modified thresholding." Computers & Electrical Engineering (2012), In Press

[5] Istepanian, Robert S. H., Leontios J. Hadjileontiadis, and Stavros M. Panas. "ECG data compression using wavelets and higher order statistics methods." Information Technology in Biomedicine, IEEE Transactions on 5.2 (2001): 108-115.

[6] Sharma, L. N., S. Dandapat, and Anil Mahanta. "Multichannel ECG data compression based on multiscale principal component analysis." IEEE Transactions on Information technology in Biomedicine 16.4 (2012):730-736.

[7] Raeiatibanadkooki, Mahsa, et al. "Compression and encryption of ECG signal using wavelet and chaotically Huffman code in telemedicine application." Journal of medical systems 40.3 (2016): 1-8.

[8] Miaou, Shaou-Gang, and Shu-Nien Chao. "Wavelet-based lossy-to-lossless ECG compression in a unified vector quantization framework." Biomedical Engineering, IEEE Transactions on 52.3 (2005): 539-543.

[9] Jha, Chandan Kumar, and Maheshkumar H. Kolekar. "Performance analysis of ECG data compression using wavelet based hybrid transform method." Microwave, Optical and Communication Engineering (ICMOCE), 2015 International Conference on. IEEE, 2015.

[10] Singh, Brij N., and Arvind K. Tiwari. Optimal selection of wavelet basis function applied to ECG signal denoising." Digital Signal Processing 16.3 (2006): 275-287

[11] Rajoub, Bashar A. "An efficient coding algorithm for the compression of ECG signals using the wavelet transform." IEEE transactions on biomedical engineering 49.4 (2002): 355-362.

[12] Gordhandas, Ankit J., Thomas Heldt, and George C. Verghese. "Real-Time Extraction and Analysis of Key Morphological Features in the Electrocardiogram, for Data Compression and Clinical Decision Support." AAAI 2011 Spring Symposium on Computational Physiology (SS-11-04). AAAI. 2011, pp. 15-18

[13] Bilgin, Ali, Michael W. Marcellin, and Maria I. Altbach. "Compression of electrocardiogram signals using JPEG2000." Consumer Electronics, IEEE Transactions on 49.4 (2003): 833-840.

[14] Al-Fahoum, Amjed S. "Quality assessment of ECG compression techniques using a wavelet-based diagnostic measure." Information Technology in Biomedicine, IEEE Transactions on 10.1 (2006): 182-191.