Combined Principal Component Analysis and Compression of 12-Lead Electrocardiogram signal using Singular Value Decomposition

Dr. Mridul Kumar Mathur[1], Gyan Prakash Bissa[2]

Dept of Comp. Sc., LMCST, Jodhpur (Raj.)[1]
I.T. dept., N.L.U., Jodhpur (Raj.) [2]

Abstract*—Electrocardiogram (ECG) is the bioelectric signals generated during the cardiac cycle. ECG shows the actual functioning of the heart and various heart diseases can be diagnosed after analyzing ECG signals. In addition to that ECG signals are required to be compressed for cost-effective storage and fast transmission over low bandwidth channels to the remote locations. Principal Component Analysis identifies the most important features of ECG signals and ECG signal can be compressed using these Principal Components. Principal Components with high variance can be retained and low variance can be discarded. Thus the ECG signal can be compressed. Multiple leads of ECG signal can depict the clear picture of the heart therefore in this paper singular value decomposition (SVD) is used to obtained combined Principal Components of twelve lead Physikalisch-Technische Bundesanstalt (PTB) Data Base ECG signal. Along with this ECG signal is also compressed and reconstructed with the help of first few Eigen Vectors and quantity of compression is measured by Compression Ratio (CR) and reconstructed signal quality is measured by Percentage Root Mean Square (PRD) difference.*

***Keywords*—** *Compression Ratio (CR), Percentage Root Mean Square Difference (PRD) difference ,Eigen Value; Eigen Vector; Principal Components analysis;*

I. INTRODUCTION

A. Electrocardiogram (ECG)

Electrocardiogram depicts the electrical activity of the heart through recoding of bioelectric potentials. These bioelectric potentials are the time-varying voltage signals produced by the myocardium during the cardiac cycle [1]. Heart related abnormalities can be depicted through ECG. ECG provides the most important information for diagnosis of the heart. Experts residing at the remote places can analyze the ECG signal through high transmission channel for transmission of signal it furthermore requires ECG signal to be compressed for this purpose. Moreover, as numerous quantities of ECG data are generated every year, it becomes mandatory to employ efficient techniques for data storage and retrieval. These requirements lead us to compress the ECG signal for rapid transmission and to reduce storage requirement for huge amount of ECG data [2]. In addition this principal component of all the twelve lead is obtained simultaneously to extract all the features of ECG signal because only one or two lead is not compatible enough to get the entire feature related to heart.

B. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) [3] is a popular multivariate technique that extracts the most valuable information presented in the data in the form of variance. PCA is a technique that is useful for the compression and classification of data [4]. The purpose is to reduce the dimensionality of a data set by finding a new set of orthogonal variables, smaller than the original set of variables(principal components) , that retains most of the information of the data set.

C. Singular Value Decomposition (SVD)

Definition

Singular value decomposition (SVD) [5, 6] is a technique that decomposes a symmetric matrix into three matrixes: one is diagonal and rest is orthogonal to each other.

Let P be an m x n matrix such that the number of rows r is greater than or equal to the number of columns c. Then there exists:

(i) An m x n column orthogonal matrix P

(ii) An n x n diagonal matrix Q, with positive or zero elements, and

(iii) An n x n orthogonal matrix R such that:

$$A = PQR^T \qquad (1)$$

This is the Singular Value Decomposition (SVD) of P.

D. Principal component analysis (PCA) using SVD

Principal Components (PC) extracts the most of the variance presented in the data set. Therefore few of the principal components that are having the most of the variance can be used to represent the complete ECG signal and this leads to the compression of ECG signal. Principal components [7] can be obtained by

$$Y = \hat{R}^{T} X \quad (2)$$

Here $\hat{R}$, denote the matrix of eigenvectors sorted in the ascending order of Eigen Values and X is the original data set. Each column of Y represents principal components. First fewer columns contain maximum variations present in the data and rest can be discarded for data compression. Therefore, new data set can be represented in fewer dimensions.

II. MATERIALS AND METHODS

A. Principal Component Analysis, Compression and Reconstructions of ECG Signal using SVD

In this work Principal Component Analysis of ECG signal is done to extract most vital features of ECG signal and to compress ECG signal. This experiment is done on 12-lead PTBDB ECG signal in which ECG signal is compressed and reconstructed using few of the Eigen Vectors and Principal Components of these ECG signals are also obtained.

B. Methods of Evaluation for ECG data compression techniques

The efficiency of ECG compression techniques is described in terms of Percentage Mean Square Difference (PRD) and Compression Ratio [8] .The quality of reconstructed ECG signal can be examined with these criteria.

1) Compression Ratio (CR)

It is defined as the ratio of original data to compressed data without considering factors such as bandwidth, sampling rate, word length, reconstruction error, noise level etc. The CR is the ratio of the original file size to the compressed file size, given as follows:

$$CR = \frac{\text{Original File Size}}{\text{Compressed File Size}} \quad (3)$$

Larger the CR better is the compression.

2) Percentage Root Mean Square Difference (PRD)

PRD is measurement of percentage error. This factor evaluates the distortion regarding the difference between the original and the reconstructed signal. The factor can be represented by the following formula. The definition of the PRD is given by following equation:

$$PRD = \sqrt{\frac{\sum_{n=1}^{N}\left(x[n]-\hat{x}[n]^{2}\right)}{\sum_{n=1}^{N}\left(x[n]^{2}\right)}} X100 \quad (4)$$

where $x[n]$ and $\hat{x}[n]$ are the original and reconstructed signals, respectively; and N its length. For better reconstruction of the original signal, the values of the PRD should be very low.Thus collectively we require large CR and smaller PRD for better compression.

C. Procedure: Singular value decomposition (SVD) and PCA

To compress the ECG data through SVD and to obtain principal components of ECG Signal following steps are involved.

1) Arrange ECG data as R×C matrix X, where R represents a complete ECG cycle and C is the number of samples in each cycle.
2) Subtract row wise mean form X.
3) Apply SVD on matrix X to decompose the data set X.
4) SVD decomposes the matrix X into matrix P, matrix Q and matrix R. [Q and R represents Eigen Values and Eigen Vectors respectively].
5) Select number of eigenvectors required for reconstruction.
6) Repeat step 7 i^{th} times.
7) Set $X_{rec}(i) = X_{rec}(i)+Q(i)\times P(i)\times R^{T}(i)$ [X_{rec} is the reconstructed signal]
8) Store Q, P and R^{T} in a file A. File A is the compressed file.
9) Calculate the Principal Components (PCs).

   $PC = R^{T} \times$ [mean subtracted data]
10) Calculate compression ratio (CR) and percentage root square difference (PRD).

D. Source of ECG signal

The ECG signal used in this work is a 12- lead Physikalisch-Technische Bundesanstalt (PTB) Data Base .

III. COMPRESSION, RECONSTRUCTION AND PCA OF12-LEAD PTB DATABASE

Most informative portion of Multidimensional ECG signal can be extracted in their Principal Components. These Principal Components can be stored in compressed form and is used to reconstruct the ECG signal.

A. Formation of Input Data Pattern

Input patterns of ECG signal are formed by taking 5500 samples of 12 leads simultaneously. Each matrix that is input to the SVD is consisting of 5500 samples of all the 12 leads out of 15 leads. These all leads are considered simultaneously to find combined Principal components for all the leads so that the most of the features of all the leads can be extracted in the principal components. Therefore for the proper diagnosis of ECG signals all the leads have to be examined.

B. Singular Value Decomposition of ECG

Principal Component Analysis finds the directions in which the data has the most of the variations. To obtain Principal Components Eigen Value and Eigen Vectors are calculated using SVD. Figure 1 shows the first 4000 samples of raw ECG signal of all the 12 leads of subject 111 from PTBDB. It can be seen that every lead has different QRS morphology.

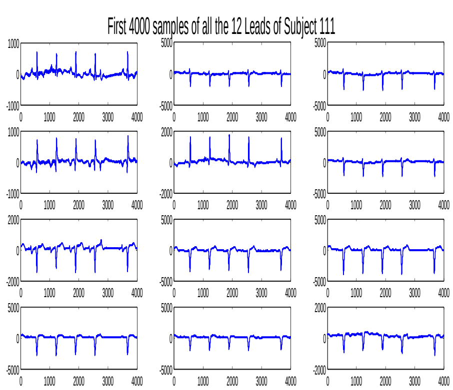

Fig 1: First 4000 samples of all the 12 Leads of the of Subject 111

Thereafter Eigen Values and Eigen Vectors of this ECG signal is obtained using the procedure explained in the section C .The ECG signal is reconstructed using 2 ,5 and 10 Eigen Vectors out of 12 Eigen Vectors which is shown in figure 2,3 and 4 respectively.

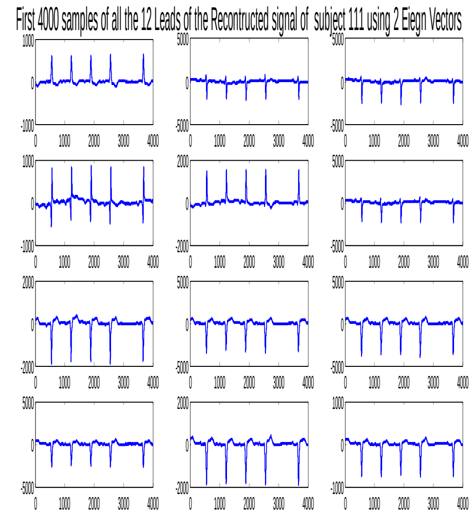

Figure 2: First 4000 samples of all the 12 leads of the Reconstructed signal of Subject 111 using 2 Eigen Vectors

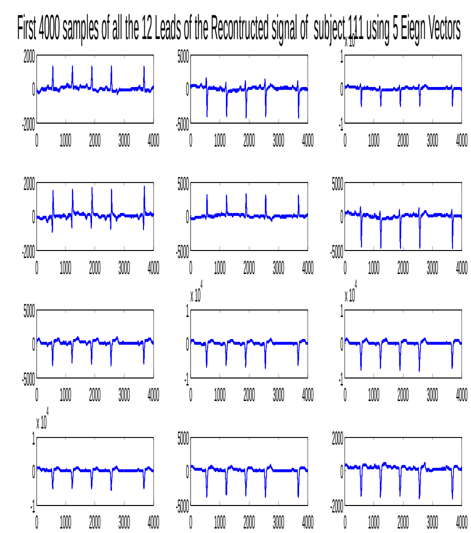

Figure 3: First 4000 samples of all the 12 leads of the Reconstructed signal of subject 111 using 5 Eigen Vectors

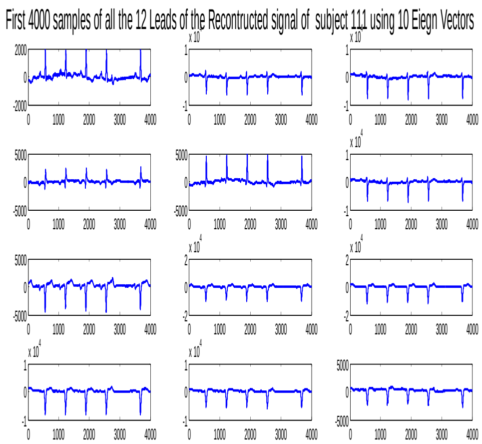

Figure 4: First 4000 samples of all the 12 leads of the Reconstructed signal of Subject 111 using 10 Eigen Vectors

			EV-2		EV-5		EV-10	
			PRD	PRD	PRD	CR (12 leads)	PRD	CR (12 leads)
PCA and SVD	111	I	23.9121	6.8977	6.8977	1.83	3.5467	1.08
PCA and SVD	002	I	23.2272	2.4423	2.4423	1.87	1.2563	1.23

Table 1: CR and PRD of subject 111 and 002 using different Eigen Vectors

It can be seen from this figure that the quality of reconstructed signal is improved. Therefore it can be said that as we increase the number of Eigen Vectors for reconstruction the quality of reconstructed signal is improved.

CR and PRD of reconstructed signal of patient 111 and 002 is shown in table 1 using 2 and 5 and 10 Eigen Vectors respectively. It implies that the quality of reconstructed signal also improves. On the contrary CR is increased as the PRD is decreased. Lowest PRD of 2.4423 is achieved for patient 111 after reconstructed using 2 Eigen Vectors.

C. Principal Component Analysis of ECG Signal

Principal components of the same signal are obtained using the procedure explained in section C. Figure 5 shows the 12 principal components of patient 111taken from PTBDB.

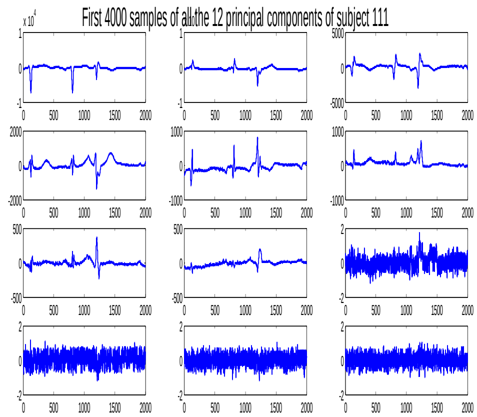

Figure 5: 12 Principal Components of subject 111

It can be seen from this figure that the first 8 principal components are showing some of the variance and rest 4 principal components are showing no variance. It implies that first 8 principal component is having some information related to this ECG signal whereas rest 4 principal components are having no information at all. Out of first 8 principal components some resemble different leads of ECG signals that are already shown in figure 1.Therefore principal components are having most of information related to a signal and rest of the principal components can be chopped off to compress the ECG signal. The raw signal, principal components and he ECG signal is reconstructed using 2 and 5 Eigen Vectors out of 10 Eigen Vectors for subject 002 which is shown in figure 6,7 ,8, 9 and 10 respectively.

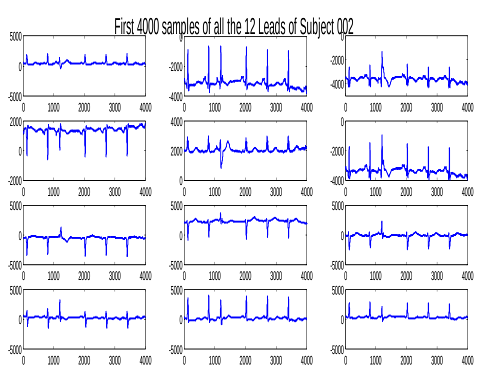

Figure 6 : First 4000 samples of all the 12 Leads of the of subject 002

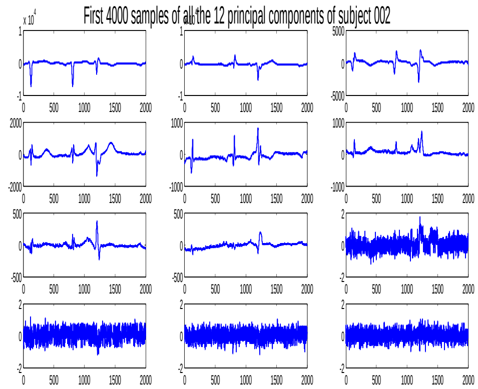

Figure 7: First 4000 samples of Principal Components of subject 002 all the 12 leads

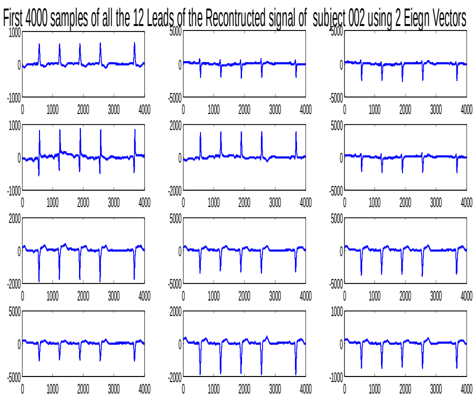

Figure 8: First 4000 samples of all the 12 leads of the Reconstructed signal of subject 002 using 2 Eigen Vectors

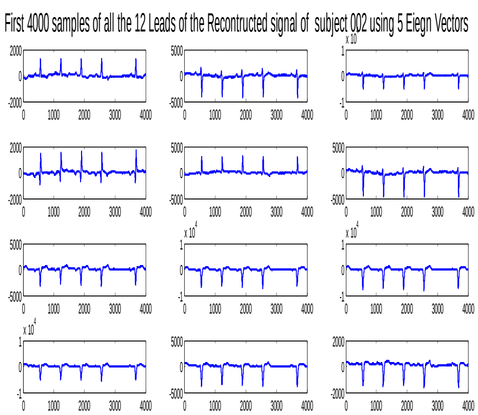

Figure 9: First 4000 samples of all the 12 leads of the Reconstructed signal of subject 002 using 5 Eigen Vectors

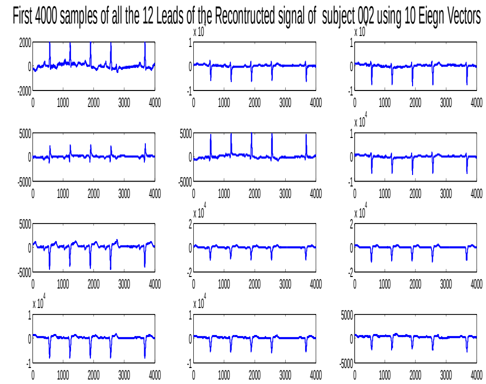

Figure 10: First 4000 samples of all the 12 leads of the Reconstructed signal of subject 002 using 10 Eigen Vectors

Similar effects and result can be seen for subject no 002.

Conclusion and Future Work

Principal Component Analysis is a tool that reduces dimension of almost all kind of information that multidimensional form and it extracts most important features of data and information with low variance can be discarded. This leads to dimension reduction or data compression.

In this work Principal component of ECG data is obtained using Singular Value Decomposition (SVD) method. This method decomposes matrix containing ECG signal into three matrixes in which one is orthogonal matrix and other two are containing Eigen Value and Eigen Vectors of ECG signal. This procedure is applied on 12-Lead Physikalisch-Technische Bundesanstalt (PTB) Data Base .

Thereafter ECG signal is reconstructed using different number of Eigen Vectors and reconstructed signals are analyzed on the basis of their PRD and CR. It is observed that as we increase number of Eigen Vectors the quality of reconstructed signal also improves. This experiment has been done for twelve lead data base. All the leads of both the subjects are successfully reconstructed.

REFERENCES

[1] Health Care in America Trends in utilization, U.S. Department of Health and Human Services, DHHS Pub No. 2004-1031

[2] National Center for Health Statistics.National Ambulatory Medical Care Survey and National Hospital Ambulatory Medical Care Survey Reason for Visit Classification and Coding Manual. Hyattsville, Maryland: Ambulatory Care Statistics Branch, Division of Health Care Statistics; 2004.

[3] *I.T, Jolliffe, Principal component analysis, New York: Springer, 2nd edition, 2002.*

[4] *H, Hotelling, Analysis of a complex of statistical variable into principal components, J.Educ. Psych., vol. 24, 417-441, 1933.*

[5] *I. A. Practical, M. Data, A. D. P. Berrar, and W. Dubitzky, "Chapter 5 Singular value decomposition and principal component analysis," pp. 1–18, 2003.*

[6] *Workalemahu, Tsegaselassie, "Singular Value Decomposition in Image Noise Filtering and Reconstruction" (2008). Mathematics Theses.Paper 52.*

[7] *R. E. Madsen, L. K. Hansen, and O. Winther, "Singular Value Decomposition and Principal Component Analysis," no.February, pp. 1–5, 2004.*

[8] *M. Blanco-Velasco, F. Cruz-Roldán, J. I. Godino-Llorente, J. Blanco-Velasco, C. Armiens-Aparicio, and F. López-Ferreras, On the Use of PRD and CR Parameters for ECG Compression, International Journal on Medical Engineering & Physics 27(6) (2005) 798-802, http://dx.doi.org/10.1016/j.medengphy.2005.02.007.*

[9] *J. Pan, W.J. Tompkins. A real-time QRS detection algorithm. IEEE Trans. Biomed.Eng. 1985; vol.32; pp. 230-236.*

[10] *Pradeepsingh, sulochana and A.K.wadhwani, "ECG data compression based on principal component analysis", vol 1, June 2013.*