



Detection of ECG points using Principal component analysis (A Review paper)

Sandeep Sharma (Scholar), Vikas Mittal (Assistant Prof.), Yuvraj Sharma (Assistant Prof.)
Deptt. Of Electronics and Communication Engg.
M.M University Mullana, Ambala, India

Abstract: *Electrocardiogram (ECG), a noninvasive technique is used as a primary diagnostic tool for cardiovascular diseases. A cleaned ECG signal provides necessary information about the electrophysiology of the heart diseases and ischemic changes that may occur. It provides valuable information about the functional aspects of the heart and cardiovascular system. The objective of paper is to automatic detection of cardiac arrhythmias in ECG signal. Recently developed digital signal processing and pattern reorganization technique is used in this thesis for detection of cardiac arrhythmias. The detection of cardiac arrhythmias in the ECG signal consists of following stages: detection of QRS complex in ECG signal; feature extraction from detected QRS complexes; classification of beat using extracted feature set from QRS complexes. In turn automatic classification of heartbeats represents the automatic detection of cardiac arrhythmias in ECG signal. Principal Component Analysis is used where lots of data, all very confusing, too many variables to consider exists, some of them are probably insignificant. PCA was invented in 1901 by Karl Pearson. It has some basic assumptions i.e. Linearity, Large variances, the principal components are orthogonal. In this paper discussing the different method of detection of ECG using PCA & various methods. The QRS complex feature is extracted based on PCA. QRS complexes feature can be presented by four largest principle components and the PCA results can be used to cluster analysis efficiently.[1]*

I. Introduction

PCA is a simple, non-parametric method for extracting relevant information from confusing data sets. PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. PCA is a special case of Factor Analysis that is highly useful in the analysis of many time series and the search for patterns of movement common to several series (true factor analysis makes different assumptions about the underlying structure and solves eigenvectors of a slightly different matrix)[2]. Cluster analysis classifies the ECG complexes based on their nature subcluster rather than supervisor-training and avoid the influence of ill training sets that often lead the classifiers based on supervisor-training fail to work.[3] Cluster analysis has been used to ECG processing, such as preprocessing training set for classifier [4], ECG character points detection [5], ECG feature parameters selection [6] and classification of ECG. The aim of this paper is to study the beat clusters and its relation with clinical categories. An integrated method for unsupervised characterization of ECG signals is developed. The approach involves PCA representation of QRS complexes in combination with self-organized clustering. With different principal components and SOM's structure, the relationship between cluster and clinical categories is revealed. PCA is a method of reducing the classic linear dimension which consists in projecting the samples on the maximum variance axes of data.[6]. This method is frequently used in detecting and locating defects and afterward in supervising the industrial and biological processes [7, 8, 9 and 10]. The PCA rests on two parts. The first step is the detection of defects which uses the PCA to model the behavior of the process in a normal state. Comparing the observed behavior with that given by the PCA, the defects are then detected.

II. Principle components analysis (PCA)

PCA in this paper is based on the Karhunen-Loeve transform. The data is described as a linear model of the form.

$$x_i = w_{qej} + e_i \quad (1)$$

$j=1$ matrix and vector notation, we can write $X=W^T + E$.

Aim is to find the optimal linear map W by minimizing the mean-squared-error (MSE) cost $J = \frac{1}{2} \|E\|^2$. For a fixed $p \times n$, the optimal subspace L_p is the one spanned by the p orthogonal eigenvectors $e_1, \dots, e_p$ which correspond to the largest eigenvalues $\lambda_1, \dots, \lambda_p$ of the covariance matrix $R_x = E(XX^T)$. The minimum MSE is the sum of the remaining eigen values. The

$$i = p+1 \quad (2)$$

Both the decoding function $X = W^T f$ and the coding function $f = UX$ are linear. Where $0 \leq j, j=1, \dots, p$ are the hidden representation parameters called factors or features. Using the obvious primary benefit of PCA arises from quantifying the importance of each dimension for describing the variability of a data set.

PCA can also be used to compress the data, i.e. by reducing the number of dimensions, without much loss of information

III. Proposed Methods

QRS complexes are extracted to make up a matrix and each column of the matrix corresponds a QRS complexes. Firstly, the matrix is centered for each row. It is mean that make each row mean value to zero. Then the matrix is normalized and PCA is used. Consequently we get the coding function $f = UX$. When only first p row of the U is used, we decrease n input dimensional space to p input dimensional space. The input vectors use the coefficient of principal components generated by PCA. Most of SOM's maps input vectors onto 2-D space. However, if output nodes are less, they will be neighbor each other. If the output nodes are too many, the clustering result may be trivial and have less meaning. In this work, SOM's maps the input vectors onto 1-D space. Different output nodes have been tested in this work. Since there are few categories of abnormal QRS complexes in one record, we select different abnormal QRS complexes from several records so that the clustering ability of the method can be studied conveniently. Five types of QRS complexes appeared frequently in the database are selected. The first 30 desired categories QRS. Training learning occurs for one vector at a time. First the network identifies the winning neuron. Then the weights of the winning neuron and its neighbor are moved closer to the input vector. The learning procedure includes two phase. Phase 1: This phase lasts for a given number of steps. The neighborhood distance starts as the maximum distance between two neurons, and decreases to the second phase neighborhood distance. The learning rate starts at an initial learning rate and decreases until it reaches its second phase rate.

Phase 2: The neighborhood distance stays at a const. The learning rate continues to decrease from initial learning rate of this phase, but very slowly. The analysis was executed after the experiments were finished and approved. The source code for ECG and Respiratory signal Principal component analysis was developed MATLAB (R2010a) (Math Works Inc.). Firstly we got data from subjects using BSL3.7.6 software with help of MP35. Secondly, we open the files (ECG & Respiratory data) in BSL PRO 3.7.6, then we got the data of ECG and Respiratory in text format. Finally we analyze the text data using the appropriate Mat-lab. source code. We got the Two PC's (principal components) of Respiratory and ECG signal plot for the subject those are participating in the experiment. The analysis of acquired data Table (1) is done through simply the mathematical expressions such as Eigen values and Eigen vectors.

The detection of singularities of a signal is a basic operation because these points often correspond to the important events of the signal. These moments can be determined by the wavelet transform thanks to local maxima (Maxima of the wavelet transform). The decrease in $W(u, s)$ can in fact be controlled by values of its local maxima. The term 'maximum module' is used to describe the points such as $W(u, s)$ being locally maximum. This implies that:

$$\frac{\partial W(u, s)}{\partial u} = 0 \quad (3)$$

singularities are located on the abscissa where the maxima modules of the fine-scale wavelet coefficients converge. To understand the properties of these maxima,

we write the wavelet transform as a multi-scale differential operator. If a wavelet has exactly n nil

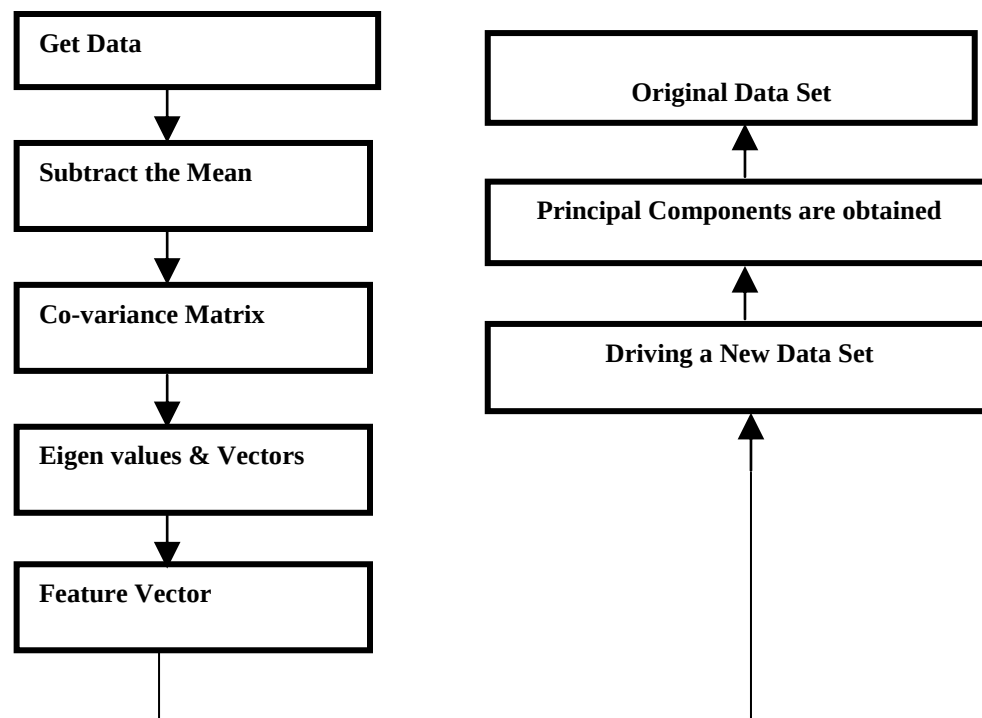


Figure.1 General Processor of Principal Component analysis

moments and a compact support is also at a compact support as

$$\psi = (-1)^n \theta^n \text{ with } \int_{-\infty}^{+\infty} \theta(t) dt \neq 0; \quad (4)$$

The wavelet transform is written as follows:

$$W_x(u, s) = s^n \frac{d^n}{du^n} (x * \bar{\theta}_s)(u) \quad (5)$$

$$\text{for } n=1, W_x(u, s) = s \frac{d}{du} (x * \bar{\theta}_s)(u) \quad (6)$$

$$\text{and for } n=2, W(u, s) = s^2 \frac{d^2}{du^2} (x * \bar{\theta})(u) \quad (7)$$

If the wavelet has only one nil moment, the maxima modules of the wavelets are the maxima of the first derivative of $x(t)$ smoothed by $s \square \square$. If the wavelet has two nil moments, the maxima module correspond to the maxima of the second derivative. If the wavelet has no local maximum of the fine scales, then $x(t)$ is locally regular. The Lipschitzian regularity is calculated starting from the decrease in the amplitude of the maximum modules determined at the scale level. We can measure this regularity by calculating the parameter:

$$\alpha_k = \log_2 |W_x(u_{k+1}, S_{k+1})| - \log_2 |W_x(u_k, s_k)| \quad (8)$$

In the dyadic case at the level of the resolution parameter becomes:

$$\alpha_j = \log_2 |W_x(n^{j+1}, 2^{j+1})| - \log_2 |W_x(n^j, 2^j)| \quad (9)$$

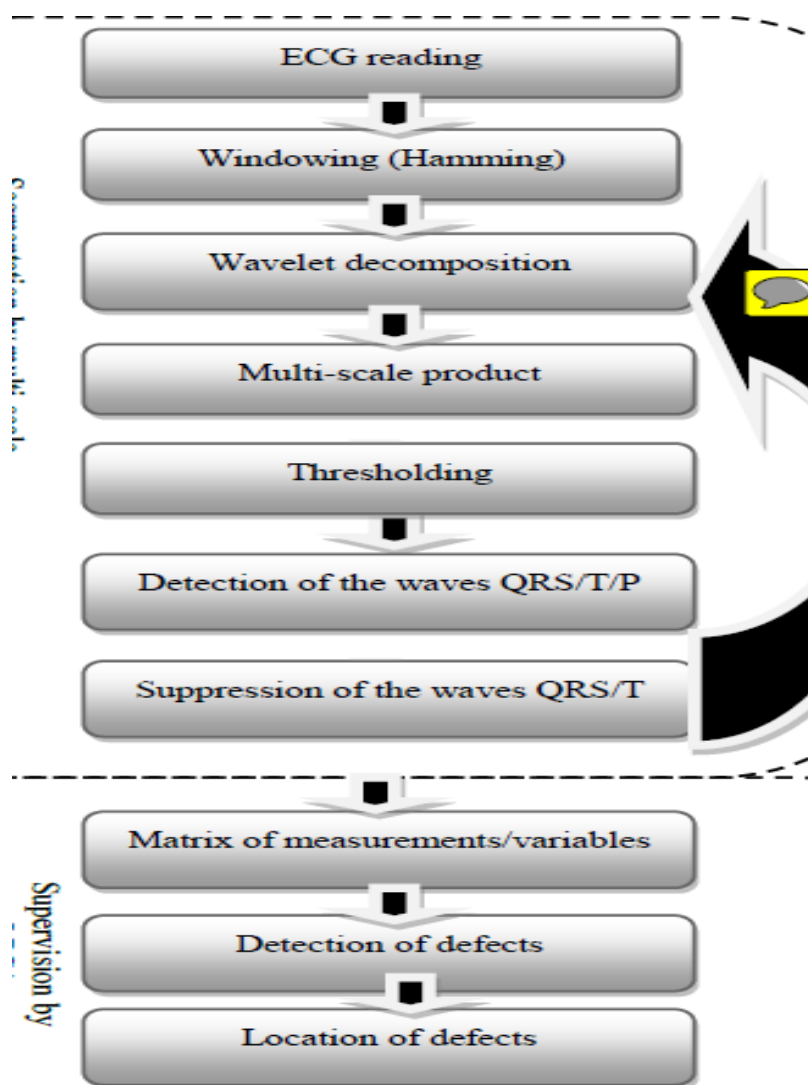


Figure: 2 Segmentation by multi-scale CWT

IV. RESULT

To test the integrated method, a set of QRS complexes is extracted from MIT-BIH. Since there are few categories of abnormal QRS complexes in one record, we select different abnormal QRS complexes from several records so that the clustering ability of the method can be studied conveniently. Five types of QRS complexes appeared frequently in the database are selected. The first 30 desired categories QRS complexes in each record are extracted. The components of the set is shown in table 1, where A is atrial premature beat, P is paced beat, V is ventricular premature beat, R is right bundle branch block and L is left bundle branch block. The extracted data of ECG complexes is centered around R peak. Considering that some PVC duration is great and sometimes R peak detection may be not the center of the complex, we have selected segment of 200ms before the fiducial point and 30ms after that

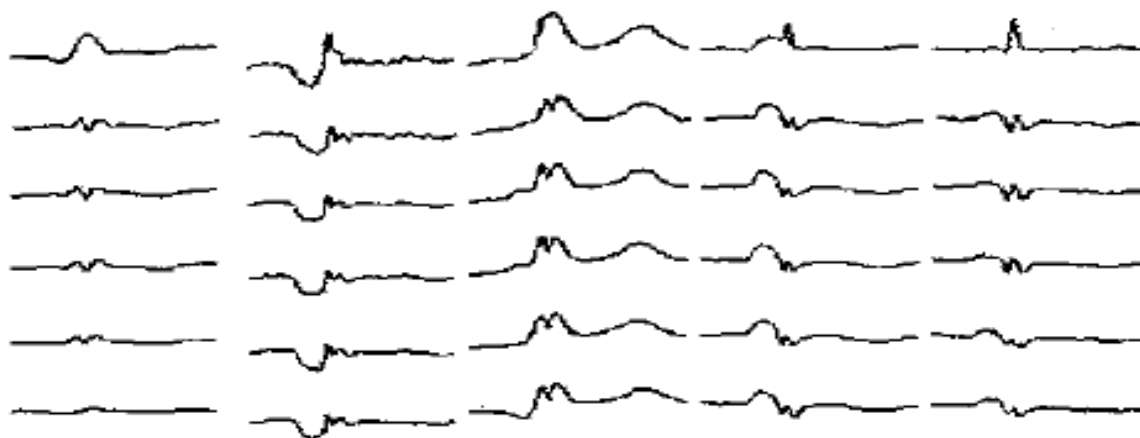


Figure:3 QRS complexes and their reconstruct with PCA

Once the ECG waves are detected (previous part), the matrix of data or measurements is determined. This matrix is composed of 500 measurements of the following variables: the amplitudes of the waves P, Q, R, S and T, and the intervals PQ, QS, ST and RR which are calculated starting from the locations of the detected waves. The choice of these variables is due to the bringing-in of information which is about the state of the ECG and the cardiovascular system afterwards. The variables introduced in this algorithm are presented as indicated on figure 5

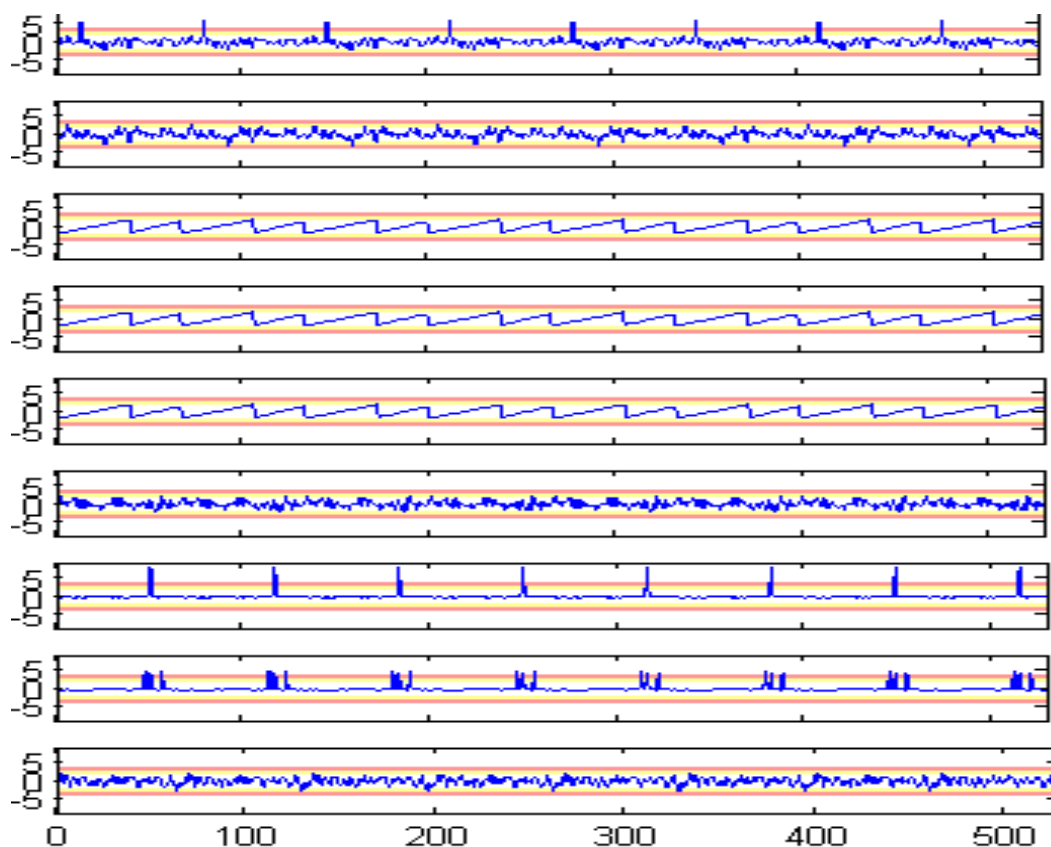


Figure:4 Progress of ECG parameters

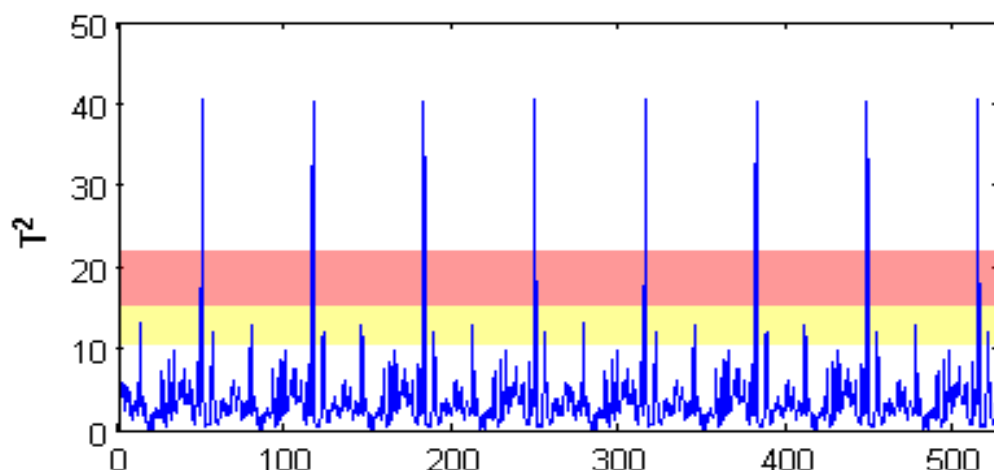


Figure: 5 Detecting defects by T

Figure 4 and 5 describe the detection of the defects by the PCA by introducing respectively the statistics T^2 and SPE. According to these figures, we notice that there is a big contradiction at the level results.

In fact, the T^2 Hotelling method detects many defects on the totality of the signal while the SPE statistic does not show any defects. Comparing these results at the real state of data (normal ECG) shows us that the SPE method is the most reliable method that is why we are going to use only this method for the location to get good results.

V. CONCLUSION

Several architectures of proposed method are considered to investigate the behavior of the proposed method. QRS Complexes feature can be presented by largest principle Components. Selecting less principle components will lead to the clustering result confusion. In addition, number of output nodes affects the cluster result. Less output nodes will also lead to confusion and too much output nodes will make the result trivial. an analysis of the ECG signal is proposed. This analysis consists of two principal steps: the segmentation and the supervision of the ECG signal. In the segmentation algorithm, we have used the multi-scale analysis of the continuous wavelet transform. This analysis is set up at three resolution levels. To detect the weak amplitude waves P and T, we need to delete the QRS complex and then the P wave. This method is evaluated by two parameters, Se and P+, which are in our case in the order of 99.9%. In fact, the segmentation of the ECG allows us to prepare the data matrix for applying the PCA to supervise this signal. It is easy to see that the first principal component is the direction along which the samples show the largest variation. The second principal component is the direction uncorrelated to the first component along which the samples show the largest variation. We have transformed our data so that is expressed in terms of the patterns between them, where the patterns closely describe the relationships between the data. We can define PCA as a meaningful graphical display of model outputs

REFERENCES

- [1] Varun Gupta., Ramveer Singh., Gavendra Singh (2011). "An Introduction to Principal Component Analysis and Its Importance in Biomedical Signal Processing" *IACSIT Press, Singapore*.
- [2] Wold, S., Esbensen, K., Geladi, P. (1987). Principal Components Analysis. *Chemometrics and Intelligent Laboratory Systems*. 2, 37-55.
- [3] Ye Wenyu, Li Gang, Lin Ling, Yu Qilian (2003). "ECG ANALYSIS BASED ON PCA AND SOM." *IEEE Int. Conf. Neural Networks 8 Signal Processing*
- [4] G.Bortolan, R. Degani, J.L. Willems, "ECG classification with neural networks and cluster analysis", *Computers in Cardiology 1991, Proceedings*, pp. 177-180, 1991
- [5] Jenn Taur Shey, Kan Ping Lin; Chang, W.H., "Measurement 12-lead ECG character pints using line segment clustering technique" *Eng. in Medicine and Biology Society, Proceedings*, vol.1, pp. 223-226, 1998
- [6] Hanen Chaouch1, Khaled Ouni2 and Lotfi Nabli (2012) "Automatic and Signal processing and Image Laboratory, National school of engineering, Monastir, Tunisia
- [7] K. Yeung, and W. Ruzzo, "Principal component analysis For clustering gene expression data," *Proceedings of the 16th International Workshop on Database and Expert Systems Applications (DEXA'05) IEEE*, 17(9), pp.763-774, 2001
- [8] B.Camille, « Détection de défauts des capteurs en ligne, validation des mesures », Université Henry Poincaré Nancy 1, Centre de Recherche Public Henri Tudor :Laboratoire de Technologie Industrielles, 2002.
- [9] J.J Sylvester "On the reduction of abilinear quantic of the nth order to the form of sum of n product products by a double orthogonal substitution", *Messenger of mathematics*, vol.19, pp.42-46, 1889

- [10] Varun Gupta, Manish Mittal, Gavendra Singh, Reena Tyagi "Time-Frequency Description of Signum Function using Principal Component & Linear Discriminant Analysis", ICCMS- 2011, Mumbai.
- [11] Zhongchao Huang "A Novel spectral Analysis Method of Atrial Fibrillation Signal Based on Hilbert-Huang Transform". Eyal Gottlieb, Sean M. Armour "Mitochondrial respiratory control is lost during growth factor deprivation."
- [12] Angel Otero "Palliative Performance Status, Heart Rate and Respiratory Rate as Predictive Factors of Survival Time in Terminally Ill Cancer Patients". Tuan Anh "Principal Components Analysis (PCA)", Financial Pricing Midterm Presentation, 2000.