

A PCA-Based Technique for QRS Complex Estimation

A Khawaja, O Dössel

Universität Karlsruhe (TH), Karlsruhe, Germany

Abstract

In this paper, a new method for QRS complex prediction is presented. It is based on Principal Components Analysis (PCA) and polynomial fitting techniques. QRS complexes were extracted from multi-lead ECG signals and were aligned very perfectly. The covariance matrix was calculated from the QRS complex data matrix of many heartbeats. Afterwards, the corresponding eigenvectors and eigenvalues were computed and the reconstruction parameters vectors were derived by expansion of every beat in terms of the first eigenvectors. Performing the first order poly-fit method on the elements of the reconstruction parameter vectors yielded certain linear functions. Thereafter, the following QRS complexes were estimated by calculating the corresponding reconstruction parameter vectors derived from these functions. The similarity, absolute error and RMS error between the original and predicted QRS complexes were measured.

1. Introduction

The analysis of ECG signal is extensively used as a diagnostic tool to provide information on the heart function. QRS complex is considered as the most striking waveform within the Electrocardiogram.

The morphology, amplitude and time of occurrence of QRS complex provide much information about the current contractile activity and state of the ventricles.

In this paper a novel method for estimating future QRS complexes of patients from the existing ones in their ECG signals using the method of Principle Components Analysis is presented.

64- and 32-channel electrocardiography is carried out on healthy young volunteers at the Institute of Biomedical Engineering, Universität Karlsruhe (TH).

QRS complexes, in different beats and the same channel, are first localized by implementing an accurate wavelet-based method for ECG delineation.

Applying Discrete Wavelet Transformation (DWT) using Haar function as Wavelet prototype and analysing the first level details coefficients formulate the methodology of the delineator, figure 1.

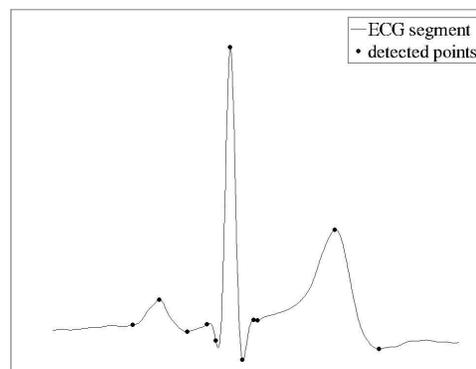


Figure 1: Lead II heart cycle derived from the 64-channel ECG along with its corresponding delineation results computed by means of our multi-lead delineation approach. The multi-channel ECG signal is recorded with 2 kHz sampling rate.

Owing to the time-varying morphology of the QRS as well as the corruption of the delineator due to noise and artefacts from various sources, filtering is an essential pre-processing stage before delineation. A new robust DWT-based approach has been devised here to segregate the baseline wander and low-frequency disturbances from ECG [1], figure 2.

Furthermore, high-frequency noise is eliminated by means of conventional Butterworth low pass filter.

After extracting QRS complexes from the ECG data set, they are further aligned by means of the cross-correlation coefficients. Each QRS is shifted toward right and left a small number of samples and finally aligned at the position corresponding to the highest cross-correlation coefficient between the QRS and a chosen template signal, the average of all QRS complexes.

The aim of QRS complex extraction and alignment is to create the input data matrix for Principal Components Analysis (PCA).

This technique is an exploratory multivariate statistical technique that allows for the identification of key variables, or combinations of variables, in a multidimensional data set that best explains the small differences between individual observations. Principal

Components Analysis is designed to capture the variance in a data set in terms of principal components [2,3].

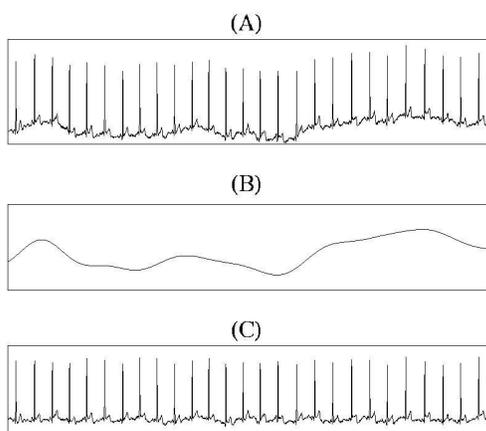


Figure 2: A) ECG signal taken from 64-channel ECG signal and distorted by baseline wander B) Estimated and extracted Baseline wander C) Filtered ECG

In other word, PCA involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible [3,4]. In this approach, each observation refers to a QRS complex extracted from multi-channel ECG data after cancelling the baseline wander and applying the delineation procedure.

2. Methods

The first step in this procedure is to build a data matrix $D = [QRS_1 \dots QRS_N]^T$ of size $N \times M$, where M is the QRS vector length and N represents the total number of heart beats in the ECG signal. This matrix is formed from the extracted and aligned QRS complexes, figure 3.

The mean QRS, named $\overline{QRS} = \frac{1}{N} \sum_{p=1}^N QRS_p$, is computed

from the data matrix. Furthermore, the covariance matrix, C (size $M \times M$), of the zero-mean data is calculated

$$C = \frac{1}{N-1} \sum_{p=1}^N (QRS_p - \overline{QRS})(QRS_p - \overline{QRS})^T$$

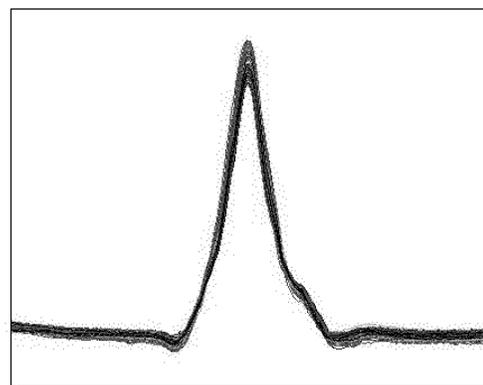


Figure 3: 332 extracted and aligned QRS complexes taken from one channel of a 64-channel ECG data

Thereafter, the eigenvectors $P_1, \dots, P_i, \dots, P_M$ and the corresponding eigenvalues $\lambda_1, \dots, \lambda_i, \dots, \lambda_M$ of the matrix C are derived and sorted in descending order. P_j is a column vector of size M and λ_j is a real number.

Denoting k , where $k \leq M$ is the number of principal components, which account together for as much of the variance in the original M variables as possible, while remaining mutually uncorrelated and orthogonal [17], the matrix \tilde{P} is defined as $\tilde{P} = [P_1 \dots P_k]$. The reconstruction parameter matrix $R = \langle b_{j,l} \rangle$ on the basis of \tilde{P} is computed as follows:

$$b_{j,l} = (P_l)^T (QRS_j - \overline{QRS}), \text{ Where } 1 \leq l \leq k \text{ and } 1 \leq j \leq N$$

The reconstruction parameter vectors, $B_1, \dots, B_j, \dots, B_N$, are the row vectors of the matrix R .

These vectors contain the variations of the shape of every heartbeat.

Using the whole set of eigenvectors, QRS complexes are reconstructed without any error: $QRS_j = \overline{QRS} + B_j P$

Choosing k number of eigenvectors, QRS complexes are reconstructed approximately: $\hat{QRS}_j = \overline{QRS} + B_j \tilde{P}$

The first element in every reconstruction parameter vector represents the coefficients of the first column vector of the matrix R .

We apply the polynomial first-order curve fitting technique on the elements of every column vectors and we get K number of functions $f_1, \dots, f_l, \dots, f_k$ [5], figure 4.

To estimate the Z^{th} QRS complex, $Z > N$, we calculate the value of former functions giving the value z as input.

The row vector $[f_1(Z), \dots, f_l(Z), \dots, f_k(Z)]$ represents the estimated reconstruction parameter vector, \hat{B}_z , of the

estimated Z^{th} QRS complex \hat{QRS}_z . The estimated QRS complex can be calculated as follows:
 $\hat{QRS}_z = \overline{QRS} + B_z \tilde{P}$

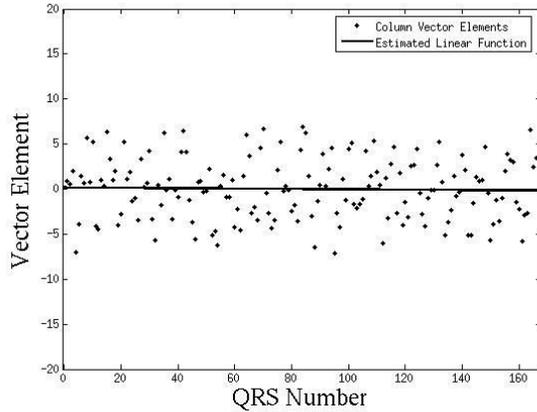


Figure 4: A plot of the l^{th} column vector elements from the matrix R along with their estimated linear function f_l versus the number of QRS complexes formed the data matrix D

To evaluate our method we compare the estimated QRS with the real corresponding QRS complex by mean of cross correlation coefficients through which common characteristics presented in the two signals can be detected. The cross correlation between two digital signals $x(n)$ and $y(n)$ having N samples each is defined as their inner or scalar product normalized by the geometric mean of their energies [7,8]:

$$\gamma_{xy} = \left(\frac{\sum_{n=1}^N x(n)y(n)}{\sqrt{\sum_{n=1}^N x^2(n) \sum_{n=1}^N y^2(n)}} \right)^{1/2}$$

γ_{xy} represents the Cross-Correlation Coefficient and n the sample. It can have a value between 0 and 1 (or 0 to 100%). Higher values of γ_{xy} signify greater similarity between the two signals and vice versa.

3. Results

The method is applied and evaluated on six 64-channel ECG signals. The duration of these databases varies between one to five minutes. Eliminating baseline wander artefacts and performing multi-channel QRS complex delineation on each signal, only half of QRS complexes from each channel are taken under consideration for PCA analysis. That is, the first half of QRS complexes in every channel formalizes the data matrix of that channel. For every matrix, its mean is

calculated, its covariance matrix is derived, and its corresponding eigenvectors and eigenvalues are computed, figure 5.

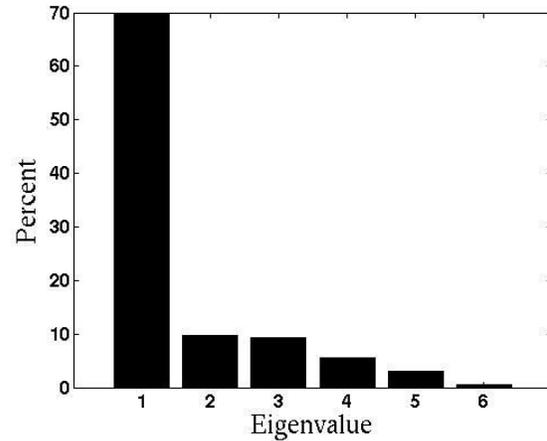


Figure 5: The percentage weight (variance) of the first eigenvalues of one QRS data matrix formed from a 64-channel ECG signal

Only k number of eigenvalues and the corresponding principal components, representing greater than 97% of all the variance, were selected. The reconstruction parameter vector for every QRS shown in the data matrix is calculated on the basis of chosen eigenvectors.

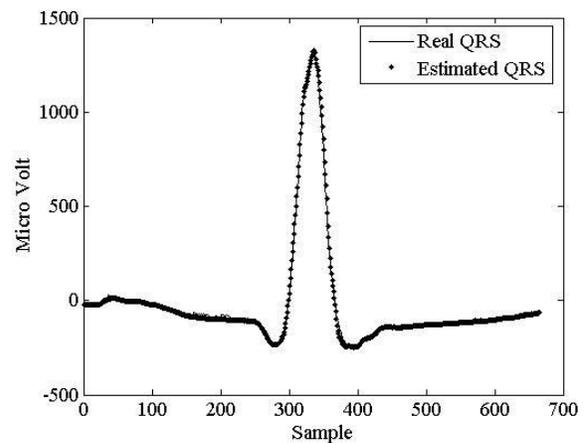


Figure 6: A plot of two QRS complexes, the first is a QRS complex in one channel of a 64-channel ECG data and the second is the corresponding estimated one yielded by our approach

Applying the polynomial fitting technique [6], k number of linear functions is derived. Giving the order

number of each QRS complex from the second half of QRS complexes to the predicted functions, the corresponding estimated reconstruction parameter vectors are yielded. Thereafter, the estimated QRS complexes are derived, figure 6. The same procedure is applied on every channel of the data set.

The similarity between each predicted and real QRS complex is derived in every channel by the mean of cross correlation coefficient. In every channel, the average value of similarities along the predicted QRS complexes is calculated, thereafter, the average similarity value for all channels is obtained.

In the same manner, the overall absolute and RMS errors between each predicted and real QRS complex of one database are computed.

Table 1: similarity values as well as error values taken for each database

Database no.	1	2	3
No. of Channel	64	64	64
No. of QRS in each Channel	332	106	57
No. of QRS Under Study	First 166	First 53	First 29
No. of Estimated QRS	166	53	28
Overall Similarity (%)	98.9742	98.5571	99.7507
Overall Absolute Error μV	16.6729	32.4993	19.7252
Overall RMS Error μV	17.5875	39.3256	19.6019

Database no.	4	5	6
No. of Channel	64	64	64
No. of QRS in each Channel	59	69	64
No. of QRS Under Study	First 30	First 35	First 32
No. of Estimated QRS	29	34	32
Overall Similarity (%)	99.641	99.3577	99.673
Overall Absolute Error μV	23.0615	16.7292	15.7831
Overall RMS Error μV	25.8759	17.4826	14.9497

The average similarity, absolute error and RMS error for all of the six databases are found to be 99.3256 %, 20.7452 μV and 22.4705 μV respectively.

4. Discussion and conclusions

The very high similarity in morphology between the estimated QRS complexes and the corresponding real signals as well as the relatively small reconstruction errors prove the efficiency, accuracy, and authenticity of this new approach. Large derivations from predicted values indicate a change in the state of the heart that might have a high diagnostics value. Analyzing the parameters and the slope of the estimation functions obtained from a short-term ECG signal (few minutes) is able to provide very useful and important information about the actual health status.

By observing these functions and setting corresponding individual thresholds for every patient, an alarm could be generated if any serious change is detected. A continuous update of this ECG analysis allows for immediate warning in case of any small change in heart status.

References

- [1] Khawaja A., Sanyal S., Dössel O. (2005) 'A Wavelet-Based Technique for Baseline Wander Correction in ECG and Multi-Channel ECG', NBC'05, IFMBE, Vol 9, 2005, p. 291-292.
- [2] Jolliffe I.T., : 'Principal Component Analysis', Springer, 2002, p. 10-76.
- [3] 'Principle Component Analysis', www.ucl.ac.uk/oncology/MicroCore/HTML_resource/PCA_1.htm.
- [4] 'Principal Component Analysis', http://www.fon.hum.uva.nl/praat/manual/Principal_component_analysis.html.
- [5] L. I. Smith. 'A Tutorial on Principal Component Analysis'. Feb 26, 2002, http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.
- [6] 'Curve Fitting Toolbox User's Guide', (The MathWork, Inc), p. 13-59.
- [7] 'Statistics Toolbox User's Guide', (The MathWork, Inc), p. 149-165.
- [8] 'Matlab Functions Reference', (The MathWork, Inc).

Address for correspondence

Antoun Khawaja
 Universität Karlsruhe, Institute of Biomedical Engineering,
 Kaiserstrasse 12, Geb. 30.33, 76131 Karlsruhe, Germany.
 Tel./Fax. : +49 (721) 608 -3851/-2789
 Antoun.Khawaja@ibt.uni-karlsruhe.de