Multiscale principal component analysis to denoise multichannel ECG signals

Sharma L. N.  Department of Electronics and Communication Engineering
Indian Institute of Technology Guwahati
Guwahati, India - 781 039
Email: lns@iitg.ernet.in

Dandapat S.  Department of Electronics and Communication Engineering
Indian Institute of Technology Guwahati
Guwahati, India - 781 039
Email: samaren@iitg.ernet.in

Mahanta A.  Department of Electronics and Communication Engineering
Indian Institute of Technology Guwahati
Guwahati, India - 781 039
Email: anilm@iitg.ernet.in

Abstract—In this work, multiscale principal component analysis (MSPCA) is introduced for denoising of multichannel electrocardiogram (MECG) signals. Wavelet decomposition of MECG signals segments the clinical information content at different Wavelet subbands or scales. At subband levels or scales multivariate data matrix are formed using Wavelet coefficients extracted from the same scales of MECG signals. At each subband matrix or scales, PCA is applied for noise elimination. To retain essential diagnostic components, matrices comprising of lower order Wavelet subbands are processed with reduced set of principal component (PC). Qualitative performance is evaluated and quantitative performance of denoising effect is measured by input/output signal-to-noise ratio (SNR). Signal distortion measures are evaluated using percentage root mean square difference (PRD), Wavelet weighted PRD (WWPRD) and Wavelet energy based diagnostic distortion measure (WEDD). The proposed algorithm is tested with database of CSE multilead measurement library. The results show significant improvement in denoising of MECG signals with the lowest PRD of 3.488 and high SNR improvement of 34.279 dB.

I. INTRODUCTION

Multichannel electrocardiogram signal processing using multivariate analysis is still remained as challenging task in biomedical engineering. It is fact that a physician needs signals having all the clinically essential diagnostic components with sufficient fidelity in an emergency department. Because this decides a critical decision.

For multivariate signal analysis, principal components analysis (PCA) is one of the oldest candidate in literature [1], [2], [3]. The variant of PCA to better extract the required information are evolving since last decade [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]. For electrocardiogram (ECG) signal enhancement, a robust extension of classical PCA is suggested [14] by analyzing shorter signal segments. PCA has been applied for data reduction, beat detection, classification, signal separation and feature extraction [15], [16]. PCA can be used for the separation of respiratory and non-respiratory segments in an ECG signal [17].

This work suggests to apply PCA at Wavelet scales for MECG signal denoising. Signals from MECG data set are Wavelet transformed. This translates the signal information in different Wavelet subbands. For ECG signal most of the diagnostic components lie in higher order subbands and lower order subbands contain high frequency information and noise. This indicates to apply PCA in lower order subband matrices. Subband matrices are formed arranging Wavelet coefficients of the same scales of all the channels of MECG data set. The selection of number of PC is based on 75% of the cumulative percentage of the total variation given by eigenvalues at each matrices. This enhance the signal while reducing the noise.

II. MULTISCALE METHOD

Multiresolution analysis of ECG signals with L level decomposition using Wavelet transform gives Lth Wavelet coefficient at jth level, w_{j,k} in ‘L + 1’ subbands. Thus we find a approximation subband at level L and details subband at level j, where j = 1, 2, ...L. For all n number of ECG channels, number of Wavelet coefficients are equal at each Wavelet scale. So, Wavelet coefficients obtained with L level decomposition of MECG signal can be arranged to form ‘L + 1’ subband matrices. In a matrix, at Wavelet scale, coefficients of similar levels of all ECG channels are placed in columns. That is, rows represent coefficients and column represent channels at a scale. So, at approximation, A_L and at details D_j are derived as multiscale matrices.

Multiscale matrices contain different parts of information from original signals due to Wavelet decomposition. These are responsible for diagnostic fidelity of the signal. To retain clinical components in the denoised signal it is essentially important to reconsider the multiscale matrices for the denoising operation. It is expected that if higher order Wavelet subband matrices treated with lower number of PC we may lose the diagnostic components.

Higher order Wavelet subbands contain more signal energy when measured in terms of energy contribution efficiency (ECE) [18],[19], [20]. In Fig.1, ECE values of all ECG leads with six level Wavelet decomposition are shown. Approximation subband is denoted as cA6 and details subbands are shown as cD1, cD2, cD3, cD4, cD5 and cD6. Standard ECG leads are shown as lead-I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5 and V6. For signals of all ECG leads higher order Wavelet subbands show more energy contribution whereas
lower order subbands show very low energy contribution. The
subbands $cD_1$, $cD_2$ & $cD_3$ contain redundant information or
noise which contribute very less to the total signal energy.
So, in proposed MSPCA based denoising for multichannel
signals, matrices consisting of these subbands are processed
with reduced set of principal components. Other matrices are
processed with all $n$ number of PCs.

After formation of multiscale matrices, $A_L$ & $D_j$, following
operations are performed for MSPCA

- Calculations of covariances and formation of covariance
  matrices from multiscale matrices
- Eigen-decomposition of covariance matrices
- Arranging eigenvalues and eigen-vectors in descended
  order for each matrix at different Wavelet scales
- Selection of PC for selected matrices
- Transformed the mean removed MECG data in PCA
  space

In this operation, the selection of PCs plays important
rule for denoising signals. Most simple rule is to compare
cumulative percentage of total variation [3] shown by selected
eigenvalues. If there are total $n$ numbers of eigenvalues and
out of these, $p$ numbers of eigenvalues are considered for
denoising, then it should satisfy

$$T_s = \frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \times 100 \geq 75\% \cdot \cdot \cdot (proposed)$$

where $\lambda_i$ is the $i$th eigenvalue and $T_s$ is the threshold set
which gives cumulative percentage of total variation explained
by number of PCs selected. The loss of signal quality depends
on the energy captured by the eigenvalues. In this work, the
threshold $T_s$ is taken equal to 75% and above which decides
the number of PC. Since the Wavelet captured most of the
ECG signal energy in higher scale matrices, the lower scale
matrices with reduced set of PCs are able to give a better
signal reconstruction.

III. RESULTS AND DISCUSSION

Standard clinical multichannel ECG signals are considered
from CSE multilead measurement library [21] data base.
Signals extracted from data set-M01-024, are subjected to
mean removal and amplitude normalization. A block of 4096
samples are taken from each channel to form a signal matrix,
$S$, of size $4096 \times 12$. Wavelet decomposition up to six
levels, using Daubechies 9/7 biorthogonal Wavelet filters, is
applied on each channel signal. Collecting Wavelet subbands
of same scale of all ECG channels, approximation matrix, $A_L$,
and details matrices, $D_j$, are formed where $L$ is the level
of decomposition and $j = 1, 2\ldots L$. For each matrices at
multiscale level PCA operation is performed using covariance
method.

Scree plot for $D_1$, $D_2$ & $D_3$ multiscale matrices is shown
in Fig.2. In Fig.2(a), eigenvalues of three multiscale matrices
$D_1$, $D_2$ & $D_3$ are shown against their corresponding principal
component. After addition of Gaussian noise with zero mean
an unity variance, the eigenvalues against PCs are plotted in
Fig.2(b). It is noticed that due to addition of Gaussian noise
the magnitude of eigenvalues are increased substantially for
lower order matrices. Also the significant number of principal
components giving larger variation of variances are increased.
This indicates that the if PCA is performed with reduced set
of PCs, it may reduce the noise. The denoising effect and loss
of information is depends on the PC selected for each Wavelet
subband matrix. The cumulative percentage of total variation
by eigenvalues, taken as 75% or more, has selected the number
of PCs for $D_1$, $D_2$ & $D_3$ as 1, 2 & 2 respectively.

In Fig.3, qualitative analysis of proposed denoising method
using MSPCA is produced and it also compare denoised
signals with conventional PCA based method. For conven-
tional PCA, the selection process for number of principal
components is kept same with proposed method. The denoised
signals of lead-I, II, III, aVR, aVL, aVF, V1, V2, V5 & V6
using proposed method show all the diagnostic components
intact. The conventional PCA has introduced distortion in all
signals and hence the 'PQRST' morphologies are severely
affected.

To demonstrate the denoising effect of the proposed method
quantitatively, the input and output SNR is measured and
shown in Table-I. A significant improvements in all the chan-
nels are noticed. The highest improvement in SNR, 34.279
dB, is seen in lead-aVR and the lowest is 12.096 dB for lead-
III.

The signal distortion measures PRD, WWPRD and WEDD
[18], [19], [20] are evaluated as existed in literatures and
shown in Table-I. For lead-V3, distortion measures are found
the lowest. The highest PRD (14.859), WWPRD (37.342) and
WEDD (7.821) are found for lead-aVL, lead-aVR and lead-I
respectively.
Eigenvales vs. principal components before addition of Gaussian noise for three multiscale matrices. Database used is CSE multilead measurement library, data set M01–024.

(a)

Eigenvales vs. principal components after addition of Gaussian noise with zero mean and unity variance. Database used is CSE multilead measurement library, data set M01–024.

(b)

Table I

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Lead-I</th>
<th>Lead-II</th>
<th>Lead-III</th>
<th>aVR</th>
<th>aVL</th>
<th>aVF</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR(O/P)</td>
<td>39.274</td>
<td>47.530</td>
<td>54.948</td>
<td>51.121</td>
<td>35.214</td>
<td>63.562</td>
<td>62.123</td>
<td>66.236</td>
<td>67.077</td>
<td>49.172</td>
<td>47.080</td>
<td>47.699</td>
</tr>
</tbody>
</table>

Distortion measures for proposed method using MSPCA

<table>
<thead>
<tr>
<th>Measures</th>
<th>Lead-I</th>
<th>Lead-II</th>
<th>Lead-III</th>
<th>aVR</th>
<th>aVL</th>
<th>aVF</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
<th>V6</th>
</tr>
</thead>
</table>

IV. Conclusion

In this work, denoising of MECG signals using MSPCA is introduced. Lower order matrices are processed with reduced set of PCs, since Wavelet translates most of the signal energy to the higher order subbands. Number of principal component selection in Wavelet multiscale matrices is based on cumulative percentage of total variation shown by eigenvalues. In this work it is taken as 75% and above. The performance of the proposed method is found satisfactory and it preserves all the clinically essential diagnostic components.

REFERENCES

Fig. 3. Original signals with Gaussian noise, denoised signals using proposed MSPCA and conventional PCA. In (a) noise added signal of lead-I, II, III, aVR, aVL, aVF, V1, V2, V5 & V6, (b) denoised signals using proposed MSPCA and in (c) denoised signals using conventional PCA. CSE multilead measurement library, data set-M01-024 is used.
