Self-adaptive model-based ECG denoising using features extracted by mean shift algorithm

Jingyu Yan a, Yan Lu a, Jia Liu b,⁎, Xinyu Wu b, Yangsheng Xu a,b

a Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong, China
b Shenzhen Institute of Advanced Technology, Shenzhen, China

1. Introduction

An electrocardiogram (ECG or EKG) is a set of recordings of electrical potential differences on human body produced by rhythmically heart activities, i.e., contraction and relaxation. As a non-invasive measurement of heart conditions via skin electrodes, ECG has become one of the most important tools to evaluate heart health and diagnose heart diseases.

However, the potential differences on body surface are so small, around 1 mV, that ECG usually has low signal-to-noise ratio (SNR), corrupted by unwanted interferences such as power line interference, electrode contact noise, motion artifacts, muscle contraction, baseline drift, ECG amplitude modulation with respiration, instrumentation noise and electrosurgical noise [1]. Although many efforts have been devoted to refining noise-free ECG signals in past years, in clinical applications, it is still lack of reliable signal processing tools to extract the weak ECG components contaminated with background noises [2], especially for abnormal heart beats where the morphologies of signals change fast and significantly.

Depending on whether a dynamic model is involved in denoising process, the existing approaches could be categorized into two classes. The model-free approaches directly focus on the obtained data and do not require a dynamical model to guide the filtering. Bandpass filters [3–5], which attempt to cancel noises out of the frequency range of ECG signal, are inadequate to handle the frequency overlap problem. Since normal ECG signal produces repetitive patterns, weighted averaging method [6] is considered as a good tool to reduce noises of normal beats but has difficulty in handling abnormal situation because of abrupt morphological change. Principle Component Analysis (PCA) and Independent Component Analysis (ICA) [7,8] have the ability to explore the correlations among multi-leads and denoise the signals by removing the uncorrelated components. However, if only single lead is available, their performances are lack of theoretic analysis. Neural Networks (NNs), as reviewed by [2,9], are also applied to refine signal based on statistical properties. Without supports from physiological background, the black-box method essentially requires the condition that heart beats possess the similar probabilistic properties from one to another. However, this condition cannot be satisfied in abnormal heart situations. In recent years, wavelet based filters have been utilized in various research and application fields, also including the denoising of ECG [10–12]. It is a suitable tool for isolating transient (non-stationary) changes in a time series by combining the time-domain and frequency-domain analysis [9].

Since the establishment of the simple but practicable dynamical model of ECG in 2003 [13], a series of studies on ECG denoising based on the dynamical model have reported improved performances compared with model-free approaches. Clifford et al.
applied nonlinear optimization to fit model parameters by minimizing the sum of squared error between model output and noisy signal [14]. Sameni et al. firstly modified the dynamical model from Cartesian coordinates to polar coordinates and then applied several Bayesian filters, including Extended Kalman Filter (EKF), Extended Kalman Smoother (EKS) and Unscented Kalman Filter (UKF), to remove noises involved in ECG signals, based on a set of two-dimensional state space functions [2]. In addition, recently, Sayadi and Shamsollahi extended the two-dimensional state space to a 17-dimensional case and enabled the estimate of much more parameters which have the ability to make the artificial ECG produced by dynamical model adaptable to different ECG signals [15]. Although the 17-dimensional EKF has the ability to estimate these parameters and achieve good SNR improvement in the context of normal ECG signals, its necessary assumption, that the morphological changes of heart beats during several cycles are small enough to assure these parameters have little variations from one beat to another, unfortunately cannot be satisfied by the ECG signals of a patient with heart disease. As shown in Fig. 1, abnormal ECG signals vary significantly from one beat to another and some waves even disappear. Sayadi and Shamsollahi [15] applied a nonlinear optimization scheme, the “lsqnonlin.m” proposed in [14], to calculate the initial parameters of dynamical model. However, in face of abnormal beats, the only once initialization for the whole signal is inadequate to achieve a good overall performance. In addition, without the support of physiological knowledge, the numerical fitting strategy has the risk of overfit, even applying Gaussian functions to simulate fake waves caused by noises.

Therefore, in this work, directly based on physiological background and morphology of ECG signals, we propose an alternative self-adaptive system to denoise the signals of both normal and abnormal ECG beats. Section 2 illustrates the principle and procedure of mean shift algorithm and proposes two strategies to enhance its application in this work. Based on the enhanced mean shift algorithm, feature extraction is implemented in Section 3. Section 4 applies these features to initialize the model of each beat and briefly introduce the filtering process. To verify the availability and efficacy of the proposed system, experiments are conducted based on MIT-BIH arrhythmia database in Section 5. Discussions and conclusions are provided in Section 6.

2. Mean shift algorithm and enhancement strategies

The basic principle of ECG dynamical model is to use a set of Gaussian functions to simulate the waves of a heart beat (i.e., a Gaussian function corresponds to a wave). To initialize the ECG dynamical model actually is to determine the location, amplitude, and width of each Gaussian function. These parameters are exactly the ECG features utilized by physicians in clinic analysis. Therefore, a good feature extraction is also helpful to establish the initial filtering model.

Location of each Gaussian function actually corresponds to location of the peak or vale of each wave, i.e., P, Q, R, S, or T. In other words, the model initialization equals to finding extremum of each wave in noisy background. Mean shift algorithm has the ability to synthesize discrete ECG samples to a continuous signal while keeping its morphology, based on an embedded Gaussian smoother, and locate extrema of the continuous signal, which are expected to correspond to peaks or vales of the waves in original ECG.

2.1. Principle and procedure

The mean shift procedure was firstly presented in 1975 by Fukunaga and Keinosuke [17] and widely used in image processing field [18]. Originally, it is a procedure for locating stationary points of a density function given discrete data sampled from that function and useful for detecting the modes of this density. However, if we treat a normalized signal sequence as probability density function, indexes of elements as uniformly sampling data from this probability density function, and values of elements as sample multiplicity (i.e. the number of times one datum has been sampled), mean shift technique also has the ability to find stationary points of the signal sequence, equally locating local maxima or minima. As the process is based on statistics framework, it implicitly involves a smoother. If we apply Gaussian function as smoother kernel, mean shift method is mathematically described as follows.

The first step is to convert the original discrete data sequence $S$, length of which is $N$, to a continuous signal $f(x)$ using $\delta$ interpolation.

$$ f(x) = \sum_{k=1}^{N} S(k) \delta(x - k) \quad \delta(x) = \begin{cases} 1 & (x = 0) \\ 0 & \text{otherwise} \end{cases} $$

(1)

Since $f(x)$ is non-derivable and completely reserves the noises involved in original data sequence, we apply Gaussian function as smoother kernel to obtain smoothed and derivable continuous signal $\hat{f}(x)$.

$$ \hat{f}(x) = f(x) \times G_\sigma = \sum_{k=1}^{N} S(k) \delta(x - k) \times G_\sigma(x, k) = \sum_{k=1}^{N} S(k) \times G_\sigma(x, k) $$

(2)

where $G_\sigma(x, \mu)$ stands for the Gaussian kernel with mean $\mu$ and standard deviation $\sigma$. Clearly, the gradient function of $\hat{f}(x)$, denoted as $g(x)$, is given by

$$ g(x) = \frac{d}{dx} \hat{f}(x) = \sum_{k=1}^{N} S(k) \frac{d}{dx} G_\sigma(x, k) = \sum_{k=1}^{N} S(k) \times \frac{e^{-((x-k)^2)/(2\sigma^2)}}{\sqrt{2\pi} \sigma} $$

(3)

$$ g(x) = \frac{1}{\sqrt{2\pi} \sigma^2} \sum_{k=1}^{N} S(k) \times k \times e^{-((x-k)^2)/(2\sigma^2)} - \sum_{k=1}^{N} S(k) \times x \times e^{-((x-k)^2)/(2\sigma^2)} $$

(4)

To locate local minima and maxima of $\hat{f}(x)$, we apply the conventional gradient-based iteration algorithm. Given iteration step $\delta p(i) > 0$, the iteration equation for local maxima is given by

$$ x(i + 1) = x(i) + \delta p(i) \times g(x(i)) $$

(5)

and that for local minima is

$$ x(i + 1) = x(i) - \delta p(i) \times g(x(i)) $$

(6)
To simplify the equation, we set \( \text{step}(i) \) as:

\[
\text{step}(i) = \sigma^3 \times \sqrt{2\pi} \sum_{k=1}^{N} S(k) \times e^{-((x(k)-k_i)^2)/(2\sigma^2)}
\]

and obviously \( \text{step}(i) > 0 \) because \( S \) is normalized. Then, Eqs. (5) and (6) are simplified respectively as:

\[
x(i + 1) = \frac{\sum_{k=1}^{N} S(k) \times k \times e^{-((x(k)-k_i)^2)/(2\sigma^2)}}{\sum_{k=1}^{N} S(k) \times e^{-((x(k)-k_i)^2)/(2\sigma^2)}}
\]

and

\[
x(i + 1) = 2x(i) - \frac{\sum_{k=1}^{N} S(k) \times k \times e^{-((x(k)-k_i)^2)/(2\sigma^2)}}{\sum_{k=1}^{N} S(k) \times e^{-((x(k)-k_i)^2)/(2\sigma^2)}}
\]

This result is intuitive: the local mean is shifted towards the region where the majority of the data resides, and the vector \( v(i) = x(i + 1) - x(i) \) is thus called mean shift vector.

Using a signal index \( k = 1, 2, \ldots, N \) as initial search point \( x(0) \), the algorithm iterates mean shift process according to Eqs. (8) and (9), until the norm of mean shift vector \( |v(i)| \) is below a threshold \( \text{StopDist} \) or the iteration number reaches the predetermined maximum \( \text{MaxIter} \). Then, the final value of \( x(i) \) is considered as the location of the local minimum/maximum point which covers the initial index \( k \), denoted as \( p(k) \).

One of the most significant advantages of mean shift algorithm is the function of an embedded smoother, implemented by Gaussian kernel in this paper, which refines the envelop of data sequence and removes the noises that may cause fake local extrema. Filter performance significantly depends on the standard deviation \( \sigma \). A suitable selection of \( \sigma \) can assure good match in configuration between clean signal and \( \hat{f}(x) \) and, meanwhile, eliminate noises involved in the original beat segment. Ideally, local minima and maxima of clean signal equal to those of \( \hat{f}(x) \).

If \( \sigma \to 0 \), Gaussian kernel reduces to \( \delta \) kernel, i.e. \( \hat{f}(x) = \sum_{k=1}^{N} S(k) \times \delta(x - k) \). Thus \( \hat{f}(x) \) has the same value with the original sequence at each sampling time and equals to 0 at others. Equally, we can say \( \hat{f}(x) \) is converted from original sequence without any smooth efficacy. In contrast, if \( \sigma \to \infty \), \( \hat{f}(x) \) becomes the sum of \( n \) lines, i.e. \( \hat{f}(x) = \sum_{k=1}^{N} S(k) \). Equally, we can say \( \hat{f}(x) \) is over smoothed completely. Fig. 2 illustrates examples of the inadequate, proper and over filtration.

### 2.2. Enhancement strategies

To enhance the performance of the mean shift algorithm, we propose two enhancement strategies to deal with the following issues: (1) the instinct problem that the algorithm underestimates the end point values of the signal sequence, and (2) how to find all the local extrema of the signal in a timesaving way.

#### 2.2.1. Segment expansion

According to the mean shift algorithm, the value of \( \hat{f}(x') \) is calculated depending on the values of neighboring elements, weighted by their distances to \( x' \). Since the data sequence has limited length, if \( x' \) is the start or terminal point, its left or right neighboring elements are all counted as 0, the smallest value in the normalized data sequence. Therefore, underestimation cannot be avoided at the end points, often resulting in unexpected fake extrema.

Segment expansion strategy is proposed to correct the underestimate problem by assigning the left-neighboring elements of the start point the same value as the start point and the right-neighboring elements of the terminal point the same value as the terminal point. In this case, the data segment is expanded as:

\[
S(1), \ldots, S(1), S(1), S(2), \ldots, S(N), S(N), \ldots, S(N)
\]

It is necessary to note that the extremum searching is confined within the original region and the expanded data are only used for

![Fig. 2. Efficacy of embedded Gaussian filter with different \( \sigma \). The upper subfigures show the continuous signals \( \hat{f}(x) \) converted from the same original data segment shown in lower subfigures. Left: Inadequate filtration \( \sigma = 5 \); middle: proper filtration \( \sigma = 15 \); right: over filtration \( \sigma = 40 \).](image)
calculating $\tilde{f}(x)$. The length of expanded region can be set as $3 \times \sigma$, which accounts for about 99.7% of the Gaussian window.

2.2.2. Time-efficient extremum detection

As discussed above, mean shift algorithm only has the ability to determine the local maximum/minimum $p(k)$ of the initial search point $k$. Since we need to obtain all the local extrema from the signal sequence, a natural way is to apply all the signal elements as initial search points and find their corresponding extrema. However, this process would be painfully time consuming, especially when we have a long signal sequence. Considering the fact that if the left and right neighbors of an element correspond to the same local extremum, the element also corresponds to this local extremum, a self-adaptive search strategy is proposed to accelerate search process.

Supposing we have gotten the local extremum $p(k)$ of the initial search point $k$, the next mean shift process will directly start from $k + \text{jump } p$, rather than $k + 1$, by jumping a large step, where $\text{jump } p = N/5$ in this work. If $p(k + \text{jump } p)$ and $p(k)$ is the same, then the next initial search point is $k + 2 \times \text{jump } p$. Otherwise, let $\text{jump } p = \text{jump } p/2$ and start the iteration from $k + \text{jump } p$ again, until $p(k + \text{jump } p)$ and $p(k)$ are the same or $\text{jump } p$ reduces to 1. Limited by calculation accuracy, if $|p(k) - p(k + \text{jump } p)| < \text{MinDist}$, we treat them as the same. $\text{MinDist} = N/10$ in this work.

3. Feature extraction

The application obstacle of the existing filtering methods to abnormal ECG signals is the significant physiological and morphological variation from beat to beat, causing failure of the dynamical model using a set of global parameters. Therefore, beat-based denoising is necessary in this case.

3.1. Beat segmentation

Based on physiological background, traditional methods extract heart beats sequentially consisting of P, Q, R, S and T waves. Although this beat segmentation method strictly corresponds to physiological states of heart activity, it is hard to determine the start and end points due to the lack of methods to accurately locate onset of P wave and offset of T wave in noisy signals. Thus, for easy implementation, we treat the data between two consecutive R waves as a beat segment. In this case, a beat segment consists of half of R, S and T waves in a heart beat and P, Q and half of R waves in next heart beat.

![Fig. 3. Beat segmentation method. A beat consists of the data in the region of two consecutive R waves, including the half of R, S and T waves in previous heart beat and the P, Q and half of R waves in next heart beat.](image)

Fig. 3. Beat segmentation method. A beat consists of the data in the region of two consecutive R waves, including the half of R, S and T waves in previous heart beat and the P, Q and half of R waves in next heart beat.

![Fig. 4. Wave sub-segmentation result for normal heart beat.](image)

Fig. 4. Wave sub-segmentation result for normal heart beat.

Since a heart beat consists of a serial of waves, if we directly apply the enhanced mean shift algorithm to detect all the extrema of a beat, it will be hard to determine the suitable value of standard deviation of embedded Gaussian smoother. In addition, the extrema searching in whole beat still is the simple mathematic calculation, lack of physiological background. Actually, each wave has its own physiological conditions. For example, since the duration of QRS complex usually is less 0.12 s, it is impossible to find their peaks or vales too far away from the time region. In other words, the location of each wave has a predetermined time region and hence limits the extreme searching process in a time period. According to the regular physiological data from [23] and by suitably increasing the durations for surplus capacity purpose, a beat segment is divided into 6 wave sub-segments, as given in Table 1. Two wave sub-segmentation results for normal and abnormal beats are illustrated in Figs. 4 and 5.

Overlap of sub-segment, aiming at making sure most normal extreme of each wave exist in relative sub-segment, may cause an extreme found twice if it locates in overlap region. Since overlap is so short that it is impossible to contain extrema of two waves, the extreme is modeled by only one Gaussian function to avoid twice calculation.

3.3. Estimation of location $l_i$

Confined in wave sub-segment, the enhanced mean shift algorithm can locate all the maxima and minima contained in its corresponding continues signal smoothed by embedded Gaussian

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave sub-segmentation strategy. $f_s$ is sampling frequency and N is beat length.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Start point</td>
</tr>
<tr>
<td>Terminal point</td>
</tr>
</tbody>
</table>
filter. The number of found extrema depends on the selection of standard deviation $\sigma$. Sometimes, for abnormal beat as shown in Fig. 5, some waves may disappear, leaving a monotone increasing or decreasing curve with only two extrema at end points. Since no more than one real extreme locates in each wave, a strategy to determine wave location robustly is proposed using a self-adaptive $\sigma$ selection.

Starting from a given small $\sigma$, in this work $\sigma = N/10$, find all the extrema in a wave sub-segment by the enhanced mean shift algorithm. Depending on the number of found extrema, estimation of location is divided into four situations.

Except the two extrema at end points:

- If no other extrema found, use position of the maximum at end point as location of R or R wave, or set location of S, T, P or Q wave as $-1$ to represent wave missing in abnormal beat;
- If only other one maximum (minimum) found, the position of the maximum (minimum) is set as wave location;
- If totally other two or three extrema found, position of the one with biggest $k$ th order difference is set as wave location, where $k$ th order difference of point $x$ is given by:

$$d f f_k(x) = |2S(x) - S(x - k) - S(x + k)|$$

and $k$ is selected depending on the typical wave duration. For R and R waves $k = 0.005 \times f_s$ for S and Q waves $k = 0.01 \times f_s$, for T wave $k = 0.07 \times f_s$ and for P wave $k = 0.05 \times f_s$;
- Otherwise, let $\sigma = \sigma + \Delta \sigma$ to increase the performance of embedded filter and repeat the extrema search process. $\Delta \sigma = N/20$ in this work.

Limited by computational accuracy, the extrema, time distance between which is less than 0.05 s, are treated as the same. If $\sigma$ reaches a upper limit $\max_x \sigma$, $\max_x = N/2$ in this work, and cannot meet the first three situation yet, the amplitude of Gaussian function is set as 0 simply because it is too noisy to provide reliable information.

3.4. Estimation of amplitude $a_i$

Amplitude of each wave is the height difference between wave extremum and the isopotential of the whole heart beat. Firstly, select $N/20$ points from whole beat segment, which have the smallest $k$ th order difference, $k = 0.02 \times f_s$ in this work. And then apply curve fitting to determine a line (the isopotential) which has the smallest total error of the selected points. Finally, set the height difference between isopotential and the average height of points in the neighboring region of wave extremum as the wave height. The length of neighboring region is set as $0.01 \times f_s$ in this paper.

3.5. Estimation of wave width $b_i$

A set of features determines a specific ECG dynamical model. The match between real beat signal and model output demonstrates the suitability of the features. If we assign the location and amplitude of each wave to related Gaussian function in model, only the width $b_i$ remains to be determined. This problem can be solved by minimizing the sum of square of difference between model output and the original signal, with respect to $b_i$. In this work, a “lsqnonlin” [14] based optimization approach is utilized to find the values of $b_i$. Details of ECG dynamical model will be described in next section, as expressed by (17).

4. Dynamical model-based nonlinear filtering

Two components must be involved in any model-based filtering problem, the filtering algorithm and system dynamical model. Considering the nonlinear ECG dynamical model, we have to adopt nonlinear filters to handle it. Fortunately, many well-studied nonlinear filters have been proposed in literature, such as EKF, EKS and DDFs. Meanwhile, ECG dynamical model, the essence of which is to simulate each wave by Gaussian function, has been proposed in recent years. Taking advantage of the features extracted by our proposed method, the model parameters can be initialized.

4.1. Nonlinear filtering

In the context of state space domain, a nonlinear filter refers to a signal-processing device that estimates the states of a dynamical system from noisy measurements, where either the system dynamics model or measurement model is a nonlinear function of the states. Since the well-known Kalman Filter (KF) is an optimal filter in the Minimum Mean Square Error (MMSE) sense to linear system [24], its nonlinear counterpart, the EKF, attracts a wide range of engineering applications in various fields. EKS, derived from Kalman Smoother (KS), is also proposed to use the information of current observations to generate better estimate of past states. The EKF and EKS are both based on first order Taylor approximations of state transition and observation equations; therefore they require the existence of derivatives which can be obtained with a reasonable effort. In addition, to a highly nonlinear system, the first order approximation may not be accurate enough to assure a satisfied performance or even convergent results. Norgaard et al. [25] proposed a new set of estimators, called DDFs, by polynomial approximations of the nonlinear transformations, including DD1 filter based on first order approximations and DD2 filter based on second order approximations.

To facilitate the presentation of the denoising technique, we briefly introduce the basic principle and theoretical framework of EKF and interested readers can find further implementation procedures of EKS and DDFs in references [26–29].

Consider a discrete-time nonlinear system with state vector $x_k$, input vector $u_k$ and observation vector $y_k$ at the time instance $k$. The dynamical state space model can be represented in the form:

$$x_{k+1} = f(x_k, u_k, W_k)$$

$$y_k = g(x_k, v_k)$$

where $f(\cdot)$ is the state evolution function and $g(\cdot)$ is the observation function. $W_k$ and $v_k$ stand for process and measurement noise vectors respectively, with associated covariance matrices $Q = E\{w_k w_k^T\}$ and $R = E\{v_k v_k^T\}$. To iterate the evolution process, an initial state vector should be estimated ahead, denoted...
by $x_0$. Then the expectation of initial estimates of states and covariance matrix are given by $\hat{x}_0 = E(x_0)$ and $P_0 = E((x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T)$.

For application of EKF, it is also necessary to derive the linearized state and observation equations respect to states and noises near a desired reference point $(\hat{x}_k, u_k, \hat{v}_k, \hat{v}_k)$:

\begin{equation}
\hat{x}_{k+1} \approx f(\hat{x}_k, u_k, \hat{\nu}_k) + A_k(x_k - \hat{x}_k) + F_k(w_k - \hat{\nu}_k) \\
y_k \approx g(\hat{x}_k, v_k) + C_k(x_k - \hat{x}_k) + G_k(v_k - \hat{v}_k)
\end{equation}

(12)

where

\begin{align*}
A_k &= \frac{\partial f(x, u, \nu)}{\partial x}|_{x=x_k, u=u_k} \\
C_k &= \frac{\partial g(x, v)}{\partial x}|_{x=x_k, v=v_k} \\
F_k &= \frac{\partial g(x, v)}{\partial w}|_{x=x_k, v=v_k} \\
G_k &= \frac{\partial g(x, v)}{\partial v}|_{x=x_k, v=v_k}
\end{align*}

(13)

\[\hat{v}_k = E(w_k) \text{ and } \hat{v}_k = E(v_k) \text{ are the expectations of input noise vector and measurement noise vector respectively.}\]

The Kalman filter has two distinct phases: prediction and update. The first one utilizes the state estimate based on the previous information to produce an estimate of states at the current time instance. In the update phase, measurement information at the current time instance is applied to refine this prediction to obtain a revised, (hopefully) more accurate, result again for the current time instance.

(1) Predict equations:

\[\hat{x}_{k+1} = f(\hat{x}_{k-1}+u_k, \hat{\nu}_k) \]

(14)

\[P_{k+1} = A_k P_{k-1} A_k^T + F_k P_k F_k^T \]

(2) Innovation equation:

\[e_k = y_k - \hat{y}_k \]

\[S_k = C_k P_{k-1} C_k^T + G_k R_k G_k^T \]

(15)

(3) Update equations:

\[K_k = P_{k-1} C_k^T S_k^{-1} \]

\[\hat{x}_{k+1} = \hat{x}_k + K_k e_k \]

\[P_{k+1} = P_{k-1} - K_k C_k P_{k-1}^{-1} \]

(16)

where $\hat{x}_{k-1}$ and $P_{k-1}$ are the prediction output based on past measurement and corresponding error covariance matrix, $e_k$ and $S_k$ is the innovation information obtained from current measurement and corresponding innovation covariance matrix, $\hat{x}_k$ and $P_k$ are the refined estimate results based on the Kalman gain $K_k$ derived from innovation information.

4.2. Dynamic model

From the principle of ECG dynamical model proposed in [13], six Gaussian functions are applied to model the six waves in a segmented beat. To avoid the variance of beat length, the index of beat segment is linearly-spaced mapped into the region $[0, 2\pi]$. Therefore, a heart beat can be represented as the summation of six Gaussian functions:

\[ECG(\theta) = \sum_{i \in \{R,S,T,P,Q,R\}} a_i \exp \left(-\frac{(\theta - \theta_i)^2}{2b_i^2}\right) \]

(17)

where $\theta \in [0, 2\pi]$, $a_i, b_i, \theta_i$ respectively are the amplitude, width and location of the $i$th wave or Gaussian function.

The initial model is parameterized by the features founded by enhanced mean shift algorithm. A small revision of location feature is needed to transfer it into $[0, 2\pi]$, $\theta_i = ((i - 1)/(N - 1)) \times 2\pi$.

Apply Euler forward difference to obtain its discrete iterated equation:

\[ECG(k+1) = \delta_\theta \times dECG(\theta)_{[\theta_0, \theta_1]} + ECG(k) \]

\[= - \sum_{i \in \{R,S,T,P,Q\}} \delta_\theta a_i(\theta(k) - \theta_i) \exp \left(-\frac{(\theta(k) - \theta_i)^2}{2b_i^2}\right) + ECG(k) \]

(18)

where $\delta_\theta = 2\pi/N$ is angular displacement per sampling time, $N$ is the beat length and $\theta(k) = k \delta_\theta$, i.e., $\theta(0)$ corresponds to start point of the beat segment and $\theta(N) = 2\pi$ corresponds to the last one.

The values of Gaussian parameters as $a_i, b_i, \theta_i$, the iterated Eq. (18) can be directly used as nonlinear ECG dynamical equation. However, the filtration performance highly depends on the accuracy of the given Gaussian parameters. Suffered from the noises to be filtrated and the variance caused by abnormal heart condition, actually it is difficult to obtain a set of satisfied estimates of parameters only at initialization stage. Therefore, the parameters are also treated as model states based on autoregressive model:

\[p_i(k) = p_i(k-1) + w_i(k) \]

(19)

where $p_i$ denotes any of the 18 Gaussian parameters.

Denoting state vector $x = [a_R, b_R, \theta_R, \ldots, a_S, b_S, \theta_S, \ldots]$, system input $u = \theta$ and system output $y = ECG$, an expanded nonlinear state space equation for beat-based ECG signal is established as below:

\[\begin{bmatrix}
a_R(k+1) \\
\vdots \\
a_S(k+1) \\
b_R(k+1) \\
\vdots \\
b_S(k+1) \\
\theta_R(k+1) \\
\vdots \\
\theta_S(k+1) \\
f_{19}(x(k), u(k), w_{19}(k))
\end{bmatrix} =
\begin{bmatrix}
a_R(k) + w_1(k) \\
\vdots \\
a_S(k) + w_S(k) \\
b_R(k) + w_R(k) \\
\vdots \\
b_S(k) + w_S(k) \\
\theta_R(k) + w_{13}(k) \\
\vdots \\
\theta_S(k) + w_{13}(k) \\
f_{19}(x(k), u(k), w_{19}(k))
\end{bmatrix} \]

(20)

where $f_{19}(x(k), u(k), w_{19}(k))$ is simply derived from Eq. (18):

\[f_{19}(x(k), u(k), w_{19}(k)) = - \sum_{i \in \{R,S,T,P,Q\}} \delta_\theta a_i(\Delta \theta_i(k)) \exp \left(-\frac{(\Delta \theta_i(k))^2}{2b_i^2}\right) + ECG(k) + w_{19}(k) \]

(21)

\[\Delta \theta_i(k) = u(k) - \theta_i(k).\]

Since the only observable state is the ECG signal $x_{19}$, the observation equation is given by:

\[y(k) = [0, \ldots, 0, 1]x(k) + v(k) \]

(22)

To apply EKF and EKS, the linearization matrices given by Eq. (13) are represented as:

\[A_k = \begin{bmatrix}
\frac{\partial f_{19}}{\partial x_{19}}(k) & \frac{\partial f_{19}}{\partial y}(k) & \frac{\partial f_{19}}{\partial \theta}(k) & \cdots & \frac{\partial f_{19}}{\partial \theta_{18}}(k) & \frac{0_{18\times1}}{1}
\end{bmatrix} \]

(23)
where

\[
\frac{\partial f_{19}}{\partial a_i(k)} = -\delta_i \frac{\Delta \theta_i(k)}{b_i(k)^2} \exp \left( -\frac{\Delta \theta_i(k)^2}{2b_i(k)^2} \right)
\]

\[
\frac{\partial f_{19}}{\partial b_i(k)} = 2\delta_i \frac{a_i(k) \Delta \theta_i(k)}{b_i(k)^3} \left[ 1 - \frac{\Delta \theta_i(k)^2}{2b_i(k)^2} \right] \exp \left( -\frac{\Delta \theta_i(k)^2}{2b_i(k)^2} \right)
\]

\[
\frac{\partial f_{19}}{\partial v_{i}(k)} = \delta_i \frac{a_i(k)}{b_i(k)^{\gamma}} \left[ 1 - \frac{\Delta \theta_i(k)^2}{2b_i(k)^2} \right] \exp \left( -\frac{\Delta \theta_i(k)^2}{2b_i(k)^2} \right)
\]

and \(i \in \{R, S, T, P, Q, R'\}\).

5. Experiments

To evaluate availability and efficacy of the proposed denoising method, experiments based on MIT-BIH Arrhythmia Database [16], which contains both normal and abnormal ECG segments, are conducted to compare performance based on various nonlinear filters under kinds of noises.

5.1. Noise generation and evaluation methodology

In the following experiments, we define the ECG records from database as clean signals. White noise and pink noise are added to clean signals to generate noisy signals with various SNRs. White noise is mathematically defined to have a flat power spectral density over all frequencies, that is, it contains equal power within a fixed bandwidth at any center frequency. Pink noise is a signal with a frequency spectrum such that the power spectral density is proportional to the reciprocal of the frequency. In order to comprehensively compare denoising performance, we applied three kinds of filters to explore better denoising results. We adopted the notion of SNR improvement to measure the filtering performance quantitatively, which is defined as the output SNR of the filter minus the input SNR (in decibels).

\[
im p[\text{dB}] = \text{SNR}_{\text{output}} - \text{SNR}_{\text{input}} = 10\log \left( \frac{\sum_i |x_c(i)|^2}{\sum_i |x_d(i)-x_c(i)|^2} \right) - 10\log \left( \frac{\sum_i |x_n(i)|^2}{\sum_i |x_d(i)-x_n(i)|^2} \right)
\]

(25)

where \(x_c\) is the clean ECG, \(x_n\) denotes the input noisy signal, and \(x_d\) represents the output denoised signal.

To some extent, filtration based on model actually is a trade-off process between model output and measurements. Error covariance matrices \(Q\) and \(R\), as the indicators of trade-off “weights”, affect the performance significantly. A large value for covariance matrix indicates measured signal involves a big portion of noise and denoising process relies on model more than measurement, while a small one means measurement contains so little interference that denoising process trends to follow measurement.

In the following context, we apply a set of values of \(P, Q\) and \(R\) covariance matrices suitable for \(-6\) dB SNR signals to the whole experiments with SNRs from 6 to \(-6\) dB. Due to the overestimation of noises contained in signals with SNRs larger than \(-6\) dB, our selection obviously makes denoising process rely on model excessively.

Two reasons explain the selection of covariance matrices. First, it is helpful to test performances in both match and mismatch situations. Actually, in practice it is usually difficult to obtain the accurate knowledge of SNR transcendentally. Secondly, since the selection of covariance matrices emphasizes model output, the accuracy of estimate of initial model becomes much more important. Thus, it highlights the performance of the proposed mean shift based initialization method.

Fig. 6. Initialization performances based on the proposed enhanced mean shift based initializer for both typical normal and abnormal heart beats under both white and pink noises. The marked regions from A to G are helpful to illustrate the properties of proposed initializer. (a) Initial model for normal beat under white noise. (b) Initial model for abnormal beat under white noise. (c) Initial model for normal beat under pink noise. (d) Initial model for abnormal beat under pink noise. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)
In the following experiments, the stop distance and maximum iteration time for mean shift algorithm are set as $\text{StopDist} = 10^{-5}$ and $\text{MaxIter} = 10^4$.

5.2. Performance on initialization

The first concerned performance is the initial model synthesized by parameters obtained from the proposed enhanced mean shift algorithm. Given these parameters, the initial model output is calculated by Eq. (17) and used to evaluate the initializer performance.

Typical results are shown in Fig. 6, subfigures of which illustrate the initial model to both normal and abnormal beats under both white and pink noises with SNR 5 dB. The typical results are the results helpful to illustrate properties of the proposed mean shift based initializer rather than the results which have good match with clean signals.

For white noises, as shown in subfigures (a) and (b), the initial models have already obtained good filtration performance except some segments marked in circles A, B, and C. The mismatch of segments A and B is due to extra sub-waves adding on P and R waves respectively. Two possible reasons may explain the occurrence of extra sub-waves. Firstly, the so called “clean” signals actually are the records from database, which may already contains some noises in their sampling process. In this case, the neglecting of sub-waves exactly proves our method has strong physiological background and has the ability to discard fake waves. Secondly, the sub-waves may be caused by abnormal physiological states of the subjects. In this case, the clean signals are so distinct from the general knowledge, based on which the ECG dynamical model established, that it is very hard to improve performance for any model-based method unless the proposed model is specially modified. In fact, only given the single beat, it is very hard to determine the extra sub-waves are produced by noises or abnormal physiological states, even for a experienced professional. A possible way to solve this problem is to check whether these sub-waves repeatedly occur. However, this work is out of our currently focused topic. Segment C is another typical mismatch caused by the so strong noises that dominate the noisy signal and suppress the morphology of clean signal.

Denoising of pink noises, as shown in subfigures (c) and (d), is a much more challenging problem for any filtration methodology since pink noises greatly change morphologies of clean signals. For example, in segments D and G, the noisy signals have been totally shifted up or down clean signals. In this situation, it is impossible to achieve good performance for any automatic denoising method or even manual method, unless other relevant conditions are provided, just as in segments E and F. Since our proposed method only detect one wave in a specified region, the fake waves in E and F, located in the region of P wave, have been suppressed by the real P wave. It strongly demonstrates the importance of physiological background knowledge in ECG denoising.

It is worthy to note that the existing optimization based initialization method has the risk to apply Gaussian functions to fit fake waves, such as those in E and F, and even the sub-waves in A and B, only if the fitting can reduce global errors. The lack of physiological knowledge causes that its performance greatly depends on the properties of noises.

In Fig. 7, the existing optimization based initialization method has the risk to apply Gaussian functions to fit fake waves, such as those in E and F, and even the sub-waves in A and B, only if the fitting can reduce global errors. The lack of physiological knowledge causes that its performance greatly depends on the properties of noises.

EKS and p-DD2 are the ones under pink noises. (a) Average SNR improvements versus input SNRs for normal beats. (b) SD of SNR improvements versus input SNRs for normal beats. (c) Average SNR improvements versus input SNRs for abnormal beats. (d) SD of SNR improvements versus input SNRs for abnormal beats. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)
5.3. Performance on nonlinear filters

DD2, EKF and EKS are applied to denoise noisy signals based on the model of each beat. For normal beat testing, we conduct the same experiments as given in [15] to set up references. However it should be noted that the evaluation method in [15] only calculates the second half of the filtered segments to avoid transient effects rather than evaluates all the data, as we do in this paper. Therefore, the evaluation given in [15] may achieve higher SNR improvements because it ignores the first half which usually has lower SNR improvement.

For the abnormal beats, we apply three typical records, which contain significant changes from beat to beat, to demonstrate the ability of proposed denoising method. Each result shown in following tables or figures is the average result of 100 times repeated testing to avoid the affection of different noise added.

Table 2 shows the SNR improvements of the input normal and abnormal signals with 5, 0 and 5 dB SNRs respectively. For a more detailed quantitative comparison, the mean and standard deviation of SNR improvements versus different input SNRs are illustrated in Fig. 7.

Since EKF and EKS apply the analytical linearization function in filtration process, it is no doubt they outperform DD2 which uses interpolation result obtained by function evaluation as linearization function. In our experiments, input SNRs are from 6 to 6 dB, a serial of comparatively low SNRs. In this case, EKS obtains better results than EKF because EKS actually denoises noisy signal \( L \) times according to the measurements at following \( L \) sampling instances. However, if the input SNR is so high, for example above 20 dB, that noisy signal has little difference with clean signal, the denoising process actually denoises the clean signal. Therefore, the calculation of SNR improvements has less meaning.

Fig. 8 illustrates the results for both normal and abnormal records under 5 dB white noise denoised by EKS. As it can be seen, the denoised signal follows the clean ECG tightly even when the morphology changes greatly and rapidly. Fig. 9 shows the

<table>
<thead>
<tr>
<th>File (beats)</th>
<th>5 dB White noise</th>
<th>0 dB White noise</th>
<th>-5 dB White noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DD2 EKF EKS</td>
<td>DD2 EKF EKS</td>
<td>DD2 EKF EKS</td>
</tr>
<tr>
<td>111 (1–55)</td>
<td>4.62 6.28 7.16</td>
<td>0.68 2.20 1.69</td>
<td>6.27 7.98 9.26</td>
</tr>
<tr>
<td>113 (1–60)</td>
<td>4.30 6.47 8.49</td>
<td>1.23 0.05 2.92</td>
<td>6.58 8.24 9.57</td>
</tr>
<tr>
<td>124 (10–70)</td>
<td>8.77 8.78 10.65</td>
<td>5.09 6.05 6.37</td>
<td>9.86 9.92 11.31</td>
</tr>
<tr>
<td>231 (10–70)</td>
<td>4.20 7.03 7.61</td>
<td>2.02 0.42 1.69</td>
<td>7.75 7.97 9.10</td>
</tr>
<tr>
<td>Mean</td>
<td>6.47 7.99 9.29</td>
<td>3.20 4.23 4.51</td>
<td>8.41 8.99 10.35</td>
</tr>
<tr>
<td>SD</td>
<td>1.68 1.21 1.43</td>
<td>1.70 2.64 1.83</td>
<td>1.30 0.89 0.97</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>File (beats)</th>
<th>5 dB White noise</th>
<th>0 dB White noise</th>
<th>-5 dB White noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DD2 EKF EKS</td>
<td>DD2 EKF EKS</td>
<td>DD2 EKF EKS</td>
</tr>
<tr>
<td>208 (1–60)</td>
<td>3.82 5.29 5.15</td>
<td>1.01 1.83 2.53</td>
<td>6.81 7.24 7.08</td>
</tr>
<tr>
<td>207 (1–60)</td>
<td>2.53 5.47 5.03</td>
<td>0.85 1.50 1.61</td>
<td>6.00 7.24 6.92</td>
</tr>
<tr>
<td>106 (100–160)</td>
<td>2.10 6.30 6.17</td>
<td>1.47 2.20 2.39</td>
<td>6.28 7.63 7.51</td>
</tr>
<tr>
<td>Mean</td>
<td>2.82 5.69 5.45</td>
<td>1.11 1.84 2.18</td>
<td>6.36 7.37 7.17</td>
</tr>
<tr>
<td>SD</td>
<td>0.90 0.54 0.63</td>
<td>0.32 0.35 0.50</td>
<td>0.41 0.23 0.31</td>
</tr>
</tbody>
</table>

**Fig. 8.** Denoising results of (a) normal record and (b) abnormal records under white noises.
denoising results under pink noise. Although the performance is obviously worse than that under white noise, the main features, such as location, height and width of each wave, have been suitably extracted from the noisy signal. The local variances do not affect the global morphology seriously.

It’s worthy to note that the fixed covariance matrices suitable for −6 dB SNR signals significantly overestimate the noise contained in the cleaner signals. A re-selection of suitable covariance matrices with smaller value, which emphasizes measurements rather than model output, is expected to improve the result.

6. Discussion and conclusion

This paper presented a novel mean shift based algorithm to initialize the ECG dynamical model parameters, resulting in good denoising performance for both normal and abnormal ECG signals under both white and pink noises. The proposed initializing algorithm first partitions each heart beat into several wave sub-segments based on physiological knowledge, then finds the extreme points from the noisy signal using the mean shift method, and finally determines the initial model parameters according to the information of the found extreme points. Experiments were carried out with 3 kinds of filters to denoise both normal and abnormal ECG records from the MIT-BIH arrhythmia database. Experiment results demonstrate that our algorithm is suitable for both normal and abnormal heart beats, especially when the signals vary in morphology from beat to beat.

For the model-based denoising techniques, the most important issues are the model structure selection and model parameters estimation. In fact, the Gaussian functions based model is not the only available choice for ECG modeling. Any model, which has the ability to map a time sequence to ECG waveform like output, could be applied as the internal model to guide the denoising process, depending on model derivability, linearization function or interpolation formula based filters may be applied in the denoising process. However, the most important feature of ECG is the beat morphology determined by heart activities. For example, it is impossible to apply a simulating system, which may produce more than 6 waves, to model a heart beat according to the beat definition in this paper. The extra model waves usually are fake ones caused by over fitting. In this sense, the model consisting of six Gaussian functions is appropriate for ECG modeling while the neural network and wavelet synthesis have difficulty in controlling morphology of model output.

Wave sub-segmentation is another important technique in ECG analysis because each wave represents a particular phase of heart activity and holds a comparative stable features such as order of appearance, wave duration and the length between of successive waves (i.e., the length of each segment). These properties allow a fixed but flexible wave sub-segmentation to both normal and abnormal beats as the method given in this paper. The study confined in each wave is much more accurate and meaningful in physiology than just treating the whole beat as a single study subject.

The model initialization method actually extract the features of ECG signal, including the localization, amplitude and width of each wave. Therefore, ECG features extracted by the proposed initializer may also be used to ECG automatic diagnosis by rule-based or classification-based intelligent system. In addition, the output of initial model has a good approaching to original ECG signal, as shown in Fig. 6(c), controlled by only 18 parameters. That is to say the original signal can be compressed by 18 values with only a small portion of information lost. A good example for ECG compression based on ECG Dynamical model is illustrated in [15].

However, the proposed denoising system can only work off-line presently due to the comparatively large time consumption in model initialization and denoising process. Averagely 0.67 s is required to process a beat sampled at 360 Hz, tested in a computer with 1.6 GHz CPU, 1 G RAM and C++ program. The study of online denoising technique will be included in future works.

Acknowledgment

This study is funded by Shenzhen Science and Technology Programme (no. SY200806300297A).

References


