A Bayesian Residual Transform for Signal Processing

Alexander Wong, *IEEE Member* and Xiao Yu Wang

arXiv:1410.0669v2 [stat.ME] 2 Jun 2015

Abstract—**Multi-scale decomposition has been an invaluable tool for the processing of physiological signals. Much focus in multi-scale decomposition for processing such signals have been based on scale-space theory and wavelet transforms. In this study, we take a different perspective on multi-scale decomposition by investigating the feasibility of utilizing a Bayesian-based method for multi-scale signal decomposition called Bayesian Residual Transform (BRT) for the purpose of physiological signal processing. In BRT, a signal is modeled as the summation of residual signals, each characterizing information from the signal at different scales. A deep cascading framework is introduced as a realization of the BRT. Signal-to-noise ratio (SNR) analysis using electrocardiography (ECG) signals was used to illustrate the feasibility of using the BRT for suppressing noise in physiological signals. Results in this study show that it is feasible to utilize the BRT for processing physiological signals for tasks such as noise suppression.**

Index Terms—**signal processing, physiological signals, multi-scale, noise suppression, electrocardiography**

I. INTRODUCTION

Physiological signals are signals that are measured from sensors that are either placed on or implanted into the body. Such physiological signals include those obtained using electromyography (EMG), electrocardiography (ECG), electroencephalography (EEG), photoplethysmography (PPG), and ballistocardiography (BCG). The processing and interpretation of such signals is challenging due to a number of different factors. For example, it is often difficult to obtain high-fidelity physiological signals due to noise, resulting in low signal-to-noise ratio (SNR). Traditionally, signal averaging and linear filters such as band-reject and band-pass filters have been used to process such physiological signals to suppress noise; however, such approaches have also been shown to result in signal degradation [1], [2]. As such,

The authors are with the Vision and Image Processing Lab, Department of Systems Design Engineering, University of Waterloo, 200 University Ave. West, Waterloo, Ontario, Canada, N2L 3G1. Tel.: +1 519 888 4567 x35342. Fax: +1 519 746 4791. E-mail: {a28wong, x18wang}@uwaterloo.ca.

more advanced methods for handling such physiological signals are desired.

Multi-scale decomposition has become an invaluable tool for the processing of physiological signals. In multi-scale decomposition, a signal is decomposed into a set of signals, each characterizing information about the original signal at a different scale. A common signal processing task that multi-scale decomposition has shown to provide significant benefits is noise suppression, based on the notion that the information pertaining to the noise component would be largely characterized by certain scales that are separate from the scales characterizing the desired signal. Much of literature in multi-scale decomposition for physiological signal processing has focused on scale-space theory [3]–[10] and wavelet transforms [11]–[19], with some investigations also conducted using methods such as empirical mode decomposition [20]–[22].

In scale-space theory [5], a signal $f(t)$ is decomposed into a single-parameter family of n signals, denoted by L, with a progressive decrease in fine scale signal information between successive scales:

$$L = \{l_j(t)|0 \leq j \leq n-1\}, \tag{1}$$

where t denotes time, j denotes scale, $l_j(t)$ is the signal at the jth scale, and $l_0(t) = f(t)$. By decomposing a signal into a set of signals with a progressive decrease in fine scale signal information between successive scales, one can then analyze signals at coarser scales without the influence of fine scale signal information such as that pertaining to noise, which is mainly characterized at the finer scales. As such, one can utilize scale space theory to suppress noise in a signal by perform scale space decomposition on the signal and then treating one of signals at a coarser scale as the noise-suppressed signal. However, there are several limitations to the use of scale space theory for physiological signal processing pertaining to noise suppression. First, noise suppression using scale space theory requires the careful selection of which scale represents the noise-suppressed signal, which can be challenging. Second, noise suppression using scale space theory does not facilitate for fine-

grained noise suppression at the individual scales, which limits its overall flexibility in striking a balance between noise suppression and signal structural preservation.

In wavelet decomposition [23], [24], a signal $f(t)$ is decomposed into a set of wavelet coefficients $c_{j,k}(t)$ obtained using a wavelet transform W:

$$c_{j,k}(t) = W_{\psi,f}(a,b)(t) \tag{2}$$

where ψ is the wavelet, $a = 2^{-j}$ is the dyadic dilation, and $b = k2^{-j}$ is the dyadic position. Wavelet transforms has a number of advantages for the purpose of physiological signal processing, particularly pertaining to noise suppression. First, as signal information at different scales are better separated in the wavelet domain (i.e., signal information at one scale is not contained in another scale), this facilitates fine-grained noise suppression at the individual scales to strike a balance between noise suppression and signal structural preservation. Second, scale selection when performing noise suppression using wavelet transforms is less critical than that for noise suppression using scale space theory, since all scales are considered in noise suppression using wavelet transforms as opposed to a single scale selection with scale space theory. One limitation worth noting pertaining to signal processing using wavelet transforms, particularly pertaining to noise suppression, is that signals processed using wavelet transforms can exhibit oscillation artifacts related to wavelet basis functions used in the wavelet transform, which is particular noticeable when dealing with low SNR scenarios. Therefore, given some of the limitations with both scale space theory and wavelet transforms when used for physiological signal processing, one is motivated to explore alternative approaches that can address these limitations.

Here, we take a different approach by exploring a Bayesian perspective to multi-scale signal decomposition. In this perspective, a signal is viewed as an amalgamation of a number of signals, each characterizing unique signal information at a different scale with different statistical characteristics. Taking such a perspective to the problem of multi-scale signal decomposition has a number of advantages. First, like the wavelet transform, since signal information at one scale is not contained in another scale, it allows us to achieve the benefits of taking better advantage of fine-grained noise suppression at the individual scales to strike a balance between noise suppression and signal structural preservation. Second, since signals are decomposed based on their statistical characteristics as opposed to a set of deterministic basis functions, signals processed using this approach would

not exhibit the types of basis-related artifacts associated with the use of wavelet transforms. Motivated by this, in this study, we investigate the feasibility of utilizing a new Bayesian-based method for multi-scale signal decomposition called Bayesian Residual Transform (BRT) for the purpose of physiological signal processing.

This paper is organized as follows. First, the methodology behind the proposed Bayesian Residual Transform is described in Section II. The experimental setup for evaluating the feasibility of using the BRT for suppressing noise in physiological signals via signal-to-noise ratio (SNR) analysis using electrocardiography (ECG) signals is described in Section III. The experimental results and discussion is presented in Section IV, and conclusions are drawn and future work discussed in Section V.

II. BAYESIAN RESIDUAL TRANSFORM

A full derivation of the proposed Bayesian Residual Transform (BRT) can be described as follows. In the BRT, a signal $f(t)$ is modeled as the summation of n residual signals, each characterizing signal information from the signal at increasingly coarse scales:

$$f(t) = \sum_{i=1}^{n} r_i(t) = f_{\sum,1}(t), \tag{3}$$

where $f_{\sum,j}(t)$ denote a signal representing the summation of all residual signals at scales $[j, n]$:

$$f_{\sum,j}(t) = \sum_{i=j}^{n} r_i(t), \tag{4}$$

and $r_i = \{r_i(t) | t \in T\}$ is a residual signal characterizing the signal information at the ith scale with different statistical characteristics. The residual signals at the lower scales contain fine-grained signal characteristics of the signal, while the residual signals at the higher scales contain coarse-grained signal characteristics of the signal (e.g., $r_1(t)$ characterizes the finest-grained signal characteristics, while $r_n(t)$ characterizes the coarsest-grained signal characteristics). As such, each residual signal contains unique information about the scale corresponding to a particular scale that the other residual signals do not contain. Therefore, the goal of the BRT (denoted by the function B) is to decompose a signal $f(t)$ into the set of n residual signals $r_1(t), r_2(t), \ldots, r_n(t)$:

$$\{r_1(t), r_2(t), \ldots, r_n(t)\} = B(f(t)). \tag{5}$$

Determining the set of residual signals characterizing the signal information at the different scales and whose sum is equal to $f(t)$ (i.e., Eq. 3) is a highly challenging
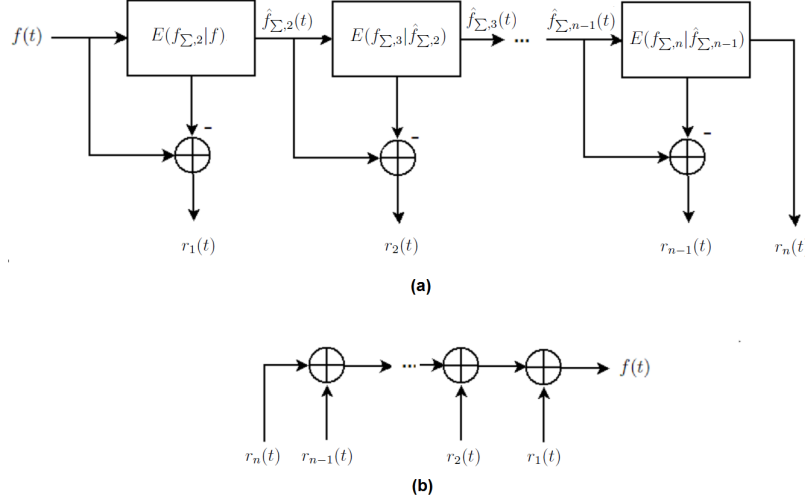
Fig. 1: Bayesian Residual Transform framework. **(a)** forward BRT. **(b)** inverse BRT.

problem, and as such with the BRT we wish to introduce a deep cascading framework to solve this problem in a more tractable manner, where a residual signal at a particular scale is computed based on computations performed at a previous scale.

Let us first rewrite Eq. 3 as follows:

$$f_{\sum,1}(t) = f_{\sum,2}(t) + r_1(t). \tag{6}$$

It can be observed from Eq. 6 that the residual signal $r_1(t)$ can be treated as the residual between the summation of all residual signals at scales $[1, n]$ and the summation of all residual signals at scales $[2, n]$. Hence, one can treat this as an inverse problem of estimating $f_{\sum,2}(t)$ given $f_{\sum,1}(t)$, with the analytical solution given by the conditional expectation $E(f_{\sum,2}(t)|f_{\sum,1}(t))$ [25] (the quantification of the conditional expectation will be explained in more detail in a later section discussing the realization of the BRT via kernel regression). Therefore, given $\hat{f}_{\sum,2}(t) = E(f_{\sum,2}(t)|f_{\sum,1}(t))$, one can substitute $\hat{f}_{\sum,2}(t)$ for $f_{\sum,2}(t)$ in Eq. 6 and rearrange the terms to obtain $r_1(t)$ as:

$$r_1(t) = f_{\sum,1}(t) - \hat{f}_{\sum,2}(t). \tag{7}$$

Given $\hat{f}_{\sum,2}(t)$, which is computed to obtain $r_1(t)$, we can express the relationship between $r_2(t)$ and $\hat{f}_{\sum,2}(t)$ in a similar manner to Eq. 6 as:

$$\hat{f}_{\sum,2}(t) = f_{\sum,3}(t) + r_2(t), \tag{8}$$

which can similarly be treated as an inverse problem of estimating $f_{\sum,3}(t)$ given $\hat{f}_{\sum,2}(t)$, with the

analytical solution given by the conditional expectation $E(f_{\sum,3}(t)|\hat{f}_{\sum,2}(t))$. Therefore, given $\hat{f}_{\sum,3}(t) = E(f_{\sum,3}(t)|\hat{f}_{\sum,2}(t))$, one can express $r_2(t)$ as:

$$r_2(t) = \hat{f}_{\sum,2}(t) - \hat{f}_{\sum,3}(t). \tag{9}$$

Generalizing this, $r_j(t)$ at scale j, for $j < n$, can be obtained by

$$r_j(t) = \hat{f}_{\sum,j}(t) - \hat{f}_{\sum,j+1}(t), \tag{10}$$

where

$$\hat{f}_{\sum,j+1}(t) = E(f_{\sum,j+1}(t)|\hat{f}_{\sum,j}(t)). \tag{11}$$

The last residual signal is computed as $r_n(t) = \hat{f}_{\sum,n}(t)$ to conform with the form expressed in Eq. 3. Hence, given Eq. 10, we have a deep cascading framework for the BRT where we can obtain the residual signal at scale j (i.e., $r_j(t)$) given the previously computed $\hat{f}_{\sum,j}(t)$. Furthermore, since the residual signal at scale $j - 1$ (i.e., $r_{j-1}(t)$) is not involved in the computation of the residual signal at scale j (i.e., $r_j(t)$) (only $\hat{f}_{\sum,j}(t)$ obtained from previous cascading step is), the information contained within $r_{j-1}(t)$ is not contained within $r_j(t)$. As such, as scale j increases, the signal information contained in $\hat{f}_{\sum,j}(t)$ becomes coarser and coarser, which results in residual signals $r_j(t)$ characterizing coarser and coarser signal information as scale increases. Based on Eq. 10, the deep cascading framework for the forward Bayesian Residual Transform (BRT) is illustrated in Fig. 1a.

Due to the condition of the summation of residual signals at all scales being equal to signal $f(t)$ (Eq. 3), the inverse BRT is simply the summation of all residual signals $r_1(t), r_2(t), \ldots, r_n(t)$:

$$f(t) = B^{-1}(r_1(t), r_2(t), \ldots, r_n(t)) = \sum_{i=1}^{n} r_i(t). \quad (12)$$

The inverse Bayesian Residual Transform (inverse BRT) procedure is illustrated in Fig. 1b.

A. Realization of Bayesian Residual Transform via Kernel Regression

In this study, we implement a realization of the BRT using a kernel regression strategy, which can be described as follows. At each iteration j, we compute $E(f_{\sum,j+1}(t)|\hat{f}_{\sum,j}(t))$ (Eq. 11) based on nonparametric Nadaraya-Watson kernel regression [26], [27] using a kernel function K_j. Here, we employ the following Gaussian kernel function K_j:

$$K_j(\hat{f}_{\sum,j}(t) - \hat{f}_{\sum,j}(t_i)) = e^{-\frac{1}{\lambda_j^2}(\hat{f}_{\sum,j}(t) - \hat{f}_{\sum,j}(t_i))^2} \quad (13)$$

Finally, the residual signal at scale n (i.e., $r_n(t)$) can be set as $E(f_{\sum,n}(t)|\hat{f}_{\sum,n-1}(t))$, which is computed at the step where $r_{n-1}(t)$ is computed. By setting $r_n(t) = E(f_{\sum,n}(t)|\hat{f}_{\sum,n-1}(t))$, the condition of the summation of signal decompositions at all scales being equal to signal $f(t)$ (i.e., Eq. 3) is satisfied. A step-by-step summary of the realization of BRT via kernel regression is shown in Algorithm 1. A step-by-step summary of the inverse BRT is shown in Algorithm 2.

B. Noise suppression

In this study, we wish to illustrate the feasibility of utilizing the BRT for processing physiological signals through the task of noise suppression. As such, we first establish a simple approach to noise suppression of signals using the BRT for illustrative purposes. The noise suppression method chosen for this study is based around the idea that the observed noisy signal $f(t)$ is formed as a summation of the desired noise-free signal $f'(t)$ and an additive noise source. Suppose that we have the true noise-free signal $f'(t)$ and we decompose it using the BRT into a series of residual signals $r_1(t), r_2(t), \ldots, r_n(t)$, where each of the residual signals characterize only information from the noise-free signal at a particular scale. Much of the information at each scale that characterizes the noise-free signal $f'(t)$ would be concentrated within only a few of the locations in each of the residual signals. What this means is that much of the information content related to $f'(t)$ is primarily concentrated within just a few locations at each scale. If we were to decompose the noisy signal

$f(t)$ using the BRT in a similar fashion, the locations of the residual signal at each scale that would otherwise have negligible information content associated with $f'(t)$ would now have low but not negligible information content that characterizes the noise source. Motivated by this, we employ a noise thresholding strategy where we only keep information from locations with information content greater than the noise information content level θ at each scale.

Algorithm 1 Step-by-step summary for Bayesian Residual Transform via Kernel Regression

Require:
  A signal $f(t)$
  parameters initialization: $\lambda_1, \ldots, \lambda_{n-1}, n$
Ensure:
  residual signals $r_1(t), r_2(t), \ldots, r_n(t)$

  $j = 1$;
  $\hat{f}_{\sum,1}(t) = f(t)$;
  while $(j < n)$ **do**
    Compute $\hat{f}_{\sum,j+1}(t) = E(f_{\sum,j+1}(t)|\hat{f}_{\sum,j}(t))$ based on kernel regression with $K_j \leftarrow$ Eq. (13)
    Compute $r_j(t) = \hat{f}_{\sum,j}(t) - \hat{f}_{\sum,j+1}(t) \leftarrow$ Eq. (10)
    $j = j + 1$;
  end while
  $r_n(t) = \hat{f}_{\sum,j}(t)$

Motivated by this, the noise thresholding strategy employed in this study can be described as follows. We first perform the forward BRT on the signal $f(t)$ to obtain n residual signals characterizing signal information at different scales $(r_1(t), r_2(t), \ldots, r_n(t))$. Since the noise information content level at each scale is not known, we employ the seminal noise level estimation method proposed by Donoho [19] to determine the noise threshold θ at each scale, which can be described as follows. At scale j, we estimate the noise threshold θ_j at each scale j using the noise-adaptive scale estimate, which can be expressed by:

$$\theta_j = MAD(r_j)/\Phi^{-1}(3/4), \quad (14)$$

where MAD is the median absolute deviation and Φ^{-1} is the normal inverse cumulative distribution function. Based on θ_j, noise thresholding is achieved to obtain noise-suppressed residual signal $r'_j(t)$ by:

$$r'_j(t) = \begin{cases} 0 & \text{if} \quad |r_j(t)| < \quad \theta_j \\ r_j(t) & \text{if} \quad\quad\quad otherwise \end{cases}$$
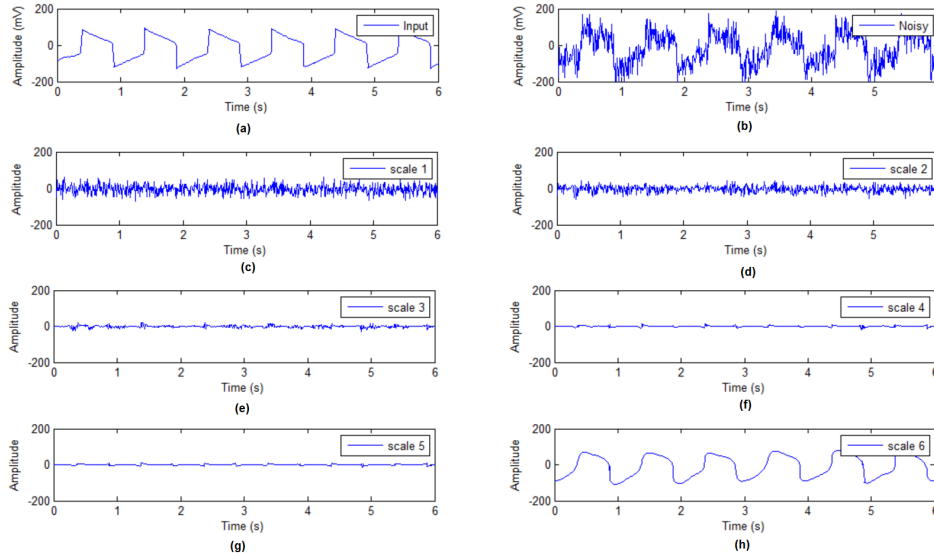
$$(15)$$

Fig. 2: Example of multi-scale signal decomposition using the BRT. **(a)** Baseline periodic test signal. **(b)** Noisy input signal with zero-mean Gaussian noise. **(c)-(h)** Signal decompositions using the BRT at different scales. It can be observed that the noise process contaminating the test signal is well characterized in the decompositions at the lower (finer) scales (scales 1 to 3), while the structural characteristics of the test signal is well characterized in the decompositions at the higher (coarser) scales (scales 4 to 6).

Finally, the inverse BRT (Eq. 12) is performed on the set of n noise-suppressed residual signals at the different scales $(r_1'(t), r_2'(t), \ldots, r_n'(t))$ to produce the noise-suppressed signal $f'(t)$. A step-by-step summary of the noise suppression method using the BRT is shown in Algorithm 3.

Algorithm 2 Step-by-step summary for inverse Bayesian Residual Transform

Require:
  residual signals $r_1(t), r_2(t), \ldots, r_n(t)$
  parameters initialization: n
Ensure:
  A signal $f(t)$

  $j = 1$;
  $f(t) = 0$;
  while $(j \leq n)$ **do**
    $f(t) = f(t) + r_j(t)$;
    $j = j + 1$;
  end while

Algorithm 3 Step-by-step summary for noise suppression using Bayesian Residual Transform

Require:
  A noisy signal $f(t)$
  parameters initialization: n
Ensure:
  A noise-suppressed signal $f'(t)$

  Perform the BRT on $f(t)$ to obtain residual signals $r_1(t), r_2(t), \ldots, r_n(t)$. ← Algorithm 1
  $j = 1$;
  while $(j \leq n)$ **do**
    Compute noise threshold θ_j ← Eq. 14;
    Compute noise-suppressed residual signal $r_j'(t)$ via threshold using θ_j ← Eq. 15;
    $j = j + 1$;
  end while
  Perform inverse BRT on $r_1'(t), r_2'(t), \ldots, r_n'(t)$ to obtain noise-suppressed signal $f'(t)$ ← Algorithm 2

III. EXPERIMENTAL SETUP

In this study, to illustrate the feasibility of utilizing the BRT for processing physiological signals, we performed a SNR analysis using electrocardiography (ECG) signals to study the performance of the BRT for the task of noise suppression. ECG signals from the MIT-BIH Normal Sinus Rhythm Database [28] were used in this study to perform the SNR analysis. This database consists of 18 ECG recordings (recorded at a sampling rate of 128 Hz) of subjects conducted at the Arrhythmia Laboratory in the Beth Israel Deaconess
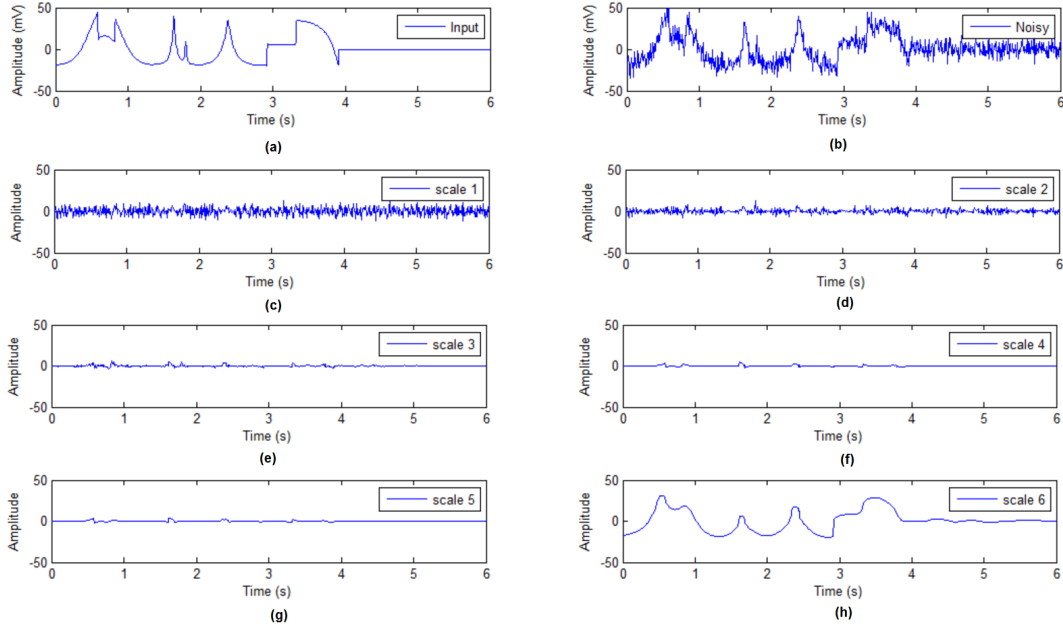
Fig. 3: Example of multi-scale signal decomposition using the BRT. **(a)** Baseline piece-wise regular test signal. **(b)** Noisy input signal with zero-mean Gaussian noise. **(c)-(h)** Signal decompositions using the BRT at different scales. It can be observed that, as with the periodic signal example, the noise process contaminating the signal is well characterized in the decompositions at the lower (finer) scales (scales 1 to 2), while the structural characteristics of the test signal is well characterized in the decompositions at the higher (coarser) scales (scales 3 to 6). Furthermore, more noticeable here than in the periodic example, it can be seen that that the decomposition at each scale exhibits good signal structural localization.

Medical Center. The subjects were found to have no significant arrhythmias. A total of 18 low-noise segments of 10 seconds was extracted, one from each recording, based on visual inspection to act as the baseline signals for evaluation. To study noise suppression performance at different SNR levels, each of the 18 baseline signals were contaminated by white Gaussian noise to produce noisy signals with SNR ranging from 12 dB to 2.5 dB (with 20 different noisy signals at each SNR), resulting in 3960 different signal perturbations used in the analysis. For comparison purposes, wavelet denoising methods with the following shrinkage rules were also used: i) Stein's Unbiased Risk (SURE) [31], ii) Heuristic SURE (HSURE) [30], iii) Universal (UNI) [30], and iv) Minimax (MINIMAX) [19]. Each of the methods uses their corresponding noise threshold and shrinkage rules in the original works. To quantitative evaluate noise suppression performance, we compute the SNR improvement as follows [32]:

$$SNRI = 10\log\left(\frac{\sum_t(f(t) - f_b(t))^2}{\sum_t(f'(t) - f_b(t))^2}\right) \quad (16)$$

where $f(t)$, $f_b(t)$, and $f'(t)$ are the noisy, baseline, and noise-suppressed signals obtained using a noise suppression method, respectively.

To study the effect of the number of scales n on noise suppression performance, the same SNR analysis is performed as described above for $n = \{2, 3, 4, 5, 6\}$.

A. Implementation details

The BRT is implemented in MATLAB (The Math-Works, Inc.), with the nonparametric conditional expectation estimates implemented in C++ and compiled as a dynamically linked MATLAB Executable (MEX) to improve computational speed. The only free parameters of the implemented realization of the BRT are the standard deviations used to model the residual signals (e.g., λ), the number of scales n, and time window size, which can be adjusted by the user to find a tradeoff between noise suppression quality and computational costs. For the SNR analysis of ECG signals, λ is set equally for all scales to the standard deviation of $f(t)$ for simplicity, n is set at 6 scales, and the time window size is set to $0.1s$. For this configuration, the current implemented realization of the BRT can process a 1028-sample signal in <1 second on an Intel(R) Core(TM) i5-3317U CPU

at 1.70GHz CPU. For the wavelet-based methods tested (SURE, HSURE, UNI, and MINIMAX), as implemented in MATLAB (The MathWorks, Inc.), soft thresholding with the Coiflet3 mother wavelet at 6 scales and single level rescaling was used as it was found to provide superior results for ECG noise suppression [33]. Each of the methods uses their corresponding noise threshold and shrinkage rules as specified in the original works.

IV. EXPERIMENTAL RESULTS

To illustrate the feasibility of utilizing the BRT for processing physiological signals, such as for the task of noise suppression, we first performed the BRT on two test signals: i) a noisy periodic test signal, and ii) a noisy piece-wise regular test signal. The multi-scale signal decomposition using the BRT on a noisy periodic test signal is shown in Fig. 2. Here, a baseline test signal (Fig. 2a) is contaminated by a zero-mean Gaussian noise process to produce a noisy signal (Fig. 2b) and then decomposed using the BRT at different scales (Figs. 2c-h). It can be observed that the noise process contaminating the signal is well characterized in the decompositions at the lower (finer) scales (scales 1 to 3), while the structural characteristics of the test signal is well characterized in the decompositions at the higher (coarser) scales (scales 4 to 6).

The multi-scale signal decomposition using the BRT on a noisy piece-wise regular test signal (generated using [29]) is shown in Fig. 3. As with the previous example, a baseline test signal (Fig. 3a) is contaminated by a zero-mean Gaussian noise process to produce a noisy signal (Fig. 3b) and then decomposed using the BRT at different scales (Figs. 3c-h). It can be observed that, as with the periodic signal example, the noise process contaminating the signal is well characterized in the decompositions at the lower (finer) scales (scales 1 to 2), while the structural characteristics of the test signal is well characterized in the decompositions at the higher (coarser) scales (scales 3 to 6). Furthermore, more noticeable here than in the periodic signal example, it can be seen that that the decomposition at each scale exhibits good signal structural localization. Therefore, given the ability of the BRT to decouple the noise process from the true signal into different scales, as illustrated in both the periodic and piece-wise regular test signals, the BRT has the potential to be useful for performing noise suppression on signals while preserving inherent signal characteristics.

In this study, to illustrate the feasibility of utilizing the BRT for processing physiological signals, we introduced a simple thresholding approach to noise suppression using the BRT for illustrative purposes (**see Section II-B**).

We then performed a quantitative SNR analysis using electrocardiography (ECG) signals from the MIT-BIH Normal Sinus Rhythm Database [28] to study the performance of the BRT for the task of noise suppression, where the SNR improvement (**see Section II-B for formulation**).

A plot of the mean SNR improvement of the tested methods vs. the different input SNRs ranging from 12 dB to 2.5 dB is shown in Fig. 4a. It can be observed that the noise-suppression method using the BRT provided strong SNR improvements across all SNRs, comparable to SURE and higher than the other 3 tested methods. It can also be observed that the UNI method consistently achieved SNR improvements below 0 dB. This is primarily due to the tendency to overestimate the noise level, resulting in signal oversmoothing and thus producing a noise-suppressed signal that is less similar to the baseline signal than the actual noisy signal. It can also be observed that the SNR improvement increases as the SNR of the input noisy signal decreases, which indicates that greater benefits are obtained through the use of noise suppression methods in low signal SNR scenarios.

To study the effect of the number of scales n on noise suppression performance, a plot of the mean SNR improvement vs. the different input SNRs ranging from 12 dB to 2.5 dB for the method using the BRT with a range of different number of scales n is shown in Fig. 4b. It can be observed that a significant gain in SNR improvement exists going from $n = 2$ to $n = 3$, with smaller SNR improvement gains from $n = 3$ all the way to $n = 6$. Furthermore, it can be observed that the SNR improvement gains from increasing the number of scales become smaller and smaller as the input SNR decreases, with the SNR improvement for $n = 3$ to $n = 6$ being approximately the same when the input SNR is 2.5 dB. Therefore, this indicates that the effect of selecting the number of scales on noise suppression performance can be significant and thus a balance between SNR improvement and the computational complexity of the BRT (which grows linearly with the number of scales) is necessary, particularly given the SNR of the noisy signal.

To study the effect of the standard deviation (SD) used for λ on noise suppression performance, a plot of the mean SNR improvement vs. the different input SNRs ranging from 12 dB to 2.5 dB for the method using the BRT with a range of different multiples of SD used for λ is shown in Fig. 4c. It can be observed that a significant gain in SNR improvement exists going from $0.5SD$ to $1SD$, with a significant drop in SNR improvements going from $1SD$ to $2SD$. Furthermore, it can be observed
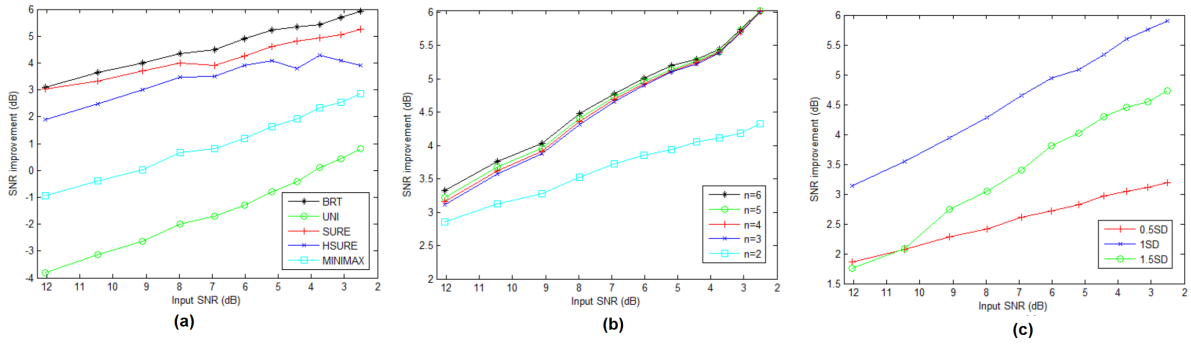
Fig. 4: Application of the BRT on ECG signals. **(a)** A plot of the mean SNR improvement vs. the different input SNRs ranging from 12 dB to 2.5 dB for the MIT-BIH Normal Sinus Rhythm Database for the tested methods. Noise-suppression method using the BRT provided strong SNR improvements across all SNRs, with performance comparable to SURE and higher than the other 3 tested methods. **(b)** A plot of the mean SNR improvement vs. the different input SNRs ranging from 12 dB to 2.5 dB for the method using the BRT with different number of scales n. **(c)** A plot of the mean SNR improvement vs. the different input SNRs ranging from 12 dB to 2.5 dB for the method using the BRT with different multiples of the standard deviation (SD) for λ.

that there are noticeable SNR improvement gains going from $0.5SD$ to $2SD$ that grows larger as the input SNR decreases. Therefore, this indicates that the effect of selecting λ on noise suppression performance can be significant, and careful selection may be important when dealing with different types of signals. For the signals tested here, it was found that $1SD$ provided the strongest results.

Typical results of noise-suppressed signals produced by the method using the BRT are shown in Fig. 5b and Fig. 5e (corresponding to two different 12 dB noisy input signals shown in Fig. 5a and Fig. 5d, respectively). Visually, it can be seen that the BRT was effectively used to produce signals with significantly reduced noise artifacts while preserving signal characteristics. Results in this study show that it is feasible to utilize the BRT for processing physiological signals for tasks such as noise suppression.

V. CONCLUSION

In this study, the feasibility of employing a Bayesian-based approach to multi-scale signal decomposition introduced here as the Bayesian Residual Transform for use in the processing of physiological signals. The Bayesian Residual Transform decomposes a signal into a set of residual signals, each characterizing information from the signal at different scales and following a particular probability distribution. This allows information at different scales to be decoupled for the purpose of signal analysis and, for the purpose of noise suppression, allows for information pertaining to the noise process contaminating the signal to be separated from the rest

of the signal characteristics. This trait is important for performing noise suppression on signals while preserving inherent signal characteristics. SNR analysis using a set of ECG signals from the MIT-BIH Normal Sinus Rhythm Database at different noise levels demonstrated that it is feasible to utilize the BRT for processing physiological signals for tasks such as noise suppression.

Given the promising results, we aim in the future to investigate alternative adaptive thresholding schemes for the task of noise suppression in physiological signals characterized by nonstationary noise, so that one can better adapt to the nonstationary noise statistics embedded at different scales. Moving beyond low-level signal processing tasks such as noise suppression, we aim with our future work to investigate and devise methods for multi-scale analysis of a signal using the Bayesian Residual Transform, which could in turn lead to improved features for signal classification. Finally, we aim to investigate the extension and generalization of the Bayesian Residual Transform for dealing with high-dimensional physiological signals such as vectorcardiographs (VCG) [34], and dealing with high-dimensional medical imaging signals from systems such as multiplexed optical high-coherence interferometry [35], optical coherence tomography [36], [37], dermatological imaging [38], diffusion weighted magnetic resonance imaging (DWI) [39]–[41], microscopy [42], [43], dynamic contrast enhanced MRI (DCE-MRI), and correlated diffusion imaging [44], [45].
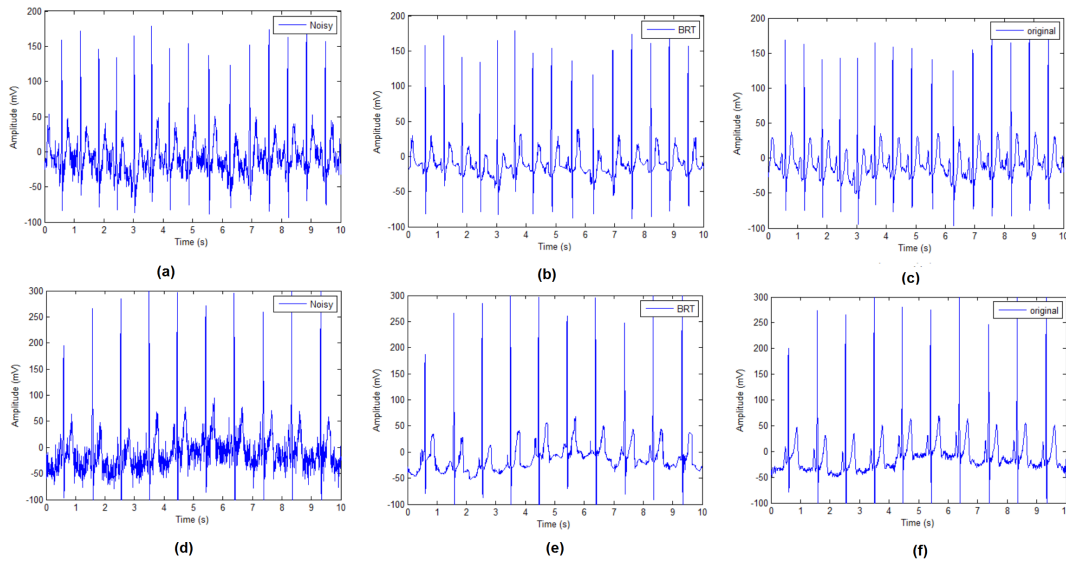
VI. ACKNOWLEDGMENT

Fig. 5: Application of the BRT on ECG signals. **(a)** Noisy input signal with SNR=12 dB, **(b)** noise-suppressed results using BRT for **a**, and **(c)** the corresponding original signal. **(d)** Another noisy input signal with SNR=12 dB, and **(e)** noise-suppressed results using BRT for **d**, and **(f)** the corresponding original signal. The results produced using the BRT has significantly reduced noise artifacts while the signal characteristics are preserved.

REFERENCES

[1] Christov, I.I. and Daskalov, I.K. Filtering of electromyogram artifacts from the electrocardiogram. *Medical Engineering and Physics* **21,** 731-736 (1999).

[2] T.D. Mewette, H. Nazeran, and K. Reynolds, Removing power line noise from recorded EMG. *Proc. IEEE Conf. Eng. Med. Bio. Society* **3,** 2190-2193 (2001).

[3] M.A. de Melo, H. Kim, and D. Nicolosi, Anisotropic median-diffusion for filtering noisy electrocardiogram signals. *Proc. IEEE Conf. Eng. Med. Bio. Society* **13,** 2562-2565 (2008).

[4] F. Jager, I. Koren, and L. Gyergyek, Multiresolution representation and analysis of ECG waveforms. *Proc. Conf. Computers. Cardiology* **1,** 547-550 (1990).

[5] Witkin, A. Scale-space filtering. *Proc. International Joint Conference on Artificial Intelligence* **2,** 1019-1022 (1983).

[6] Koenderink, J. and Van Doorn, A. The structure of images. *Biological Cybernetics* **50,** 363-370 (1984).

[7] Perona, P. and Malik, J. Scale-Space and Edge Detection Using Anisotropic Diffusion. **12,** *IEEE Trans. Pattern Analysis and Machine Intelligence* 629-639 (1990).

[8] Gilboa, G. Nonlinear scale space with spatially varying stopping time. *IEEE Trans. Pattern Analysis and Machine Intelligence* **30,** 2175-2187 (2008).

[9] A. Mishra, A. Wong, D. Clausi and P. Fieguth, "Quasi-random nonlinear scale space," Pattern Recognition Letters, vol. 31, pp. 1850-1859 (2010).

[10] A. Wong and A. Mishra, "Generalized Probabilistic Scale Space for Image Restoration," IEEE Transactions on Image Processing, vol. 19, no. 10, pp. 2774-2780 (2010).

[11] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, EMG Denoising Estimation Based on Adaptive Wavelet Thresholding for Multifunction Myoelectric Control. *Proc. CITISIA* **13,** 171-176 (2009).

[12] H.A. Kestler, M. Haschka, W. Kratz, F. Schwenker, G. Palm, V. Hombach, and M. Hoher, Denoising of high-resolution ecg-signals by combining the discrete wavelet transform with the Wiener filter. *Proc. Conf. Computers. Cardiology* **1,** 233-236 (1998).

[13] Hussain, M.S. and Mamun, M. Wavelet Denoising and Surface Electromyography Analysis. *Research J. Appl. Sciences, Eng. and Tech.* **4,** 2372-2374 (2012).

[14] Sobahi, N.M. Denoising of EMG Signals Based on Wavelet Transform. *Asian Trans. Eng.* **1,** 2372-2374 (2011).

[15] M. Popescu, P. Cristea, and A. Bezerianos, High Resolution ECG Filtering Using Adaptive Bayesian Wavelet Shrinkage. *Proc. Computers in Cardiology* **25,** 401-404 (1998).

[16] S.A. Chouakri, F. Bereksi-Reguig, S. Ahmaidi, and O. Fokapu, Wavelet denoising of the electrocardiogram signal based on the corrupted noise estimation. *Proc. Computers in Cardiology* **32,** 1021-1024 (2005).

[17] Agante, P.M. and de Sa, J.P.M. ECG noise filtering using wavelets with soft-thresholding methods. *Proc. Computers in Cardiology* **1,** 535-538 (1999).

[18] Donoho, D. and Johnstone, I.M. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81,** 425-455 (1994).

[19] Donoho, D. De-noising by soft-thresholding. *IEEE Trans. Information Theory* **41,** 613-627 (1995).

[20] Kopsinis, Y. and McLaughlin, S. Development of EMD-based Denoising Methods Inspired by Wavelet Thresholding. *IEEE Trans. Signal Processing* **57,** 1351-1362 (2009).

[21] B. Weng, M. Velasco, K. Barner, ECG denoising based on the empirical mode decomposition. *Proc. IEEE Conf. Eng. Med. Bio. Society* **1,** 1-4 (2006).

[22] M. Velasco, B. Weng, and K. Barner, ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in Biology and Medicine* **38,** 1-13 (2008).

[23] Mallat, S.G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Analysis and*

Machine Intelligence **11,** 674-693 (1989).

[24] Daubechies, I. *Ten Lectures on Wavelets* (Society for Industrial and Applied Mathematics, Philadelphia, 1992).

[25] Fieguth, P. *Statistical Image Processing and Multidimensional Modeling* (Springer, New York, 2010).

[26] Nadaraya, E. On Estimating Regression. *Theory of Probability and its Applications* **9,** 141-142 (1964).

[27] Watson, G. Smooth regression analysis. *The Indian Journal of Statistics, Series A* **26,** 359-372 (1964).

[28] A.L. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, and H. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101,** e215-e220 (2000).

[29] Peyre, G. Lecture 5 - Wavelet Compression of 1D Signals. (2006) Date of access: 20/09/2014. *https://www.ceremade.dauphine.fr/ peyre/teaching/wavelets/tp5.html*

[30] M. Misiti, Y. Misiti, G. Oppenheim, and J. Poggi, *Wavelet Toolbox Users guide* (The Mathworks, Inc., Natick, 2014).

[31] Donoho, D. and Johnstone, I. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, **90,** 1200-1224 (1995).

[32] M. Akhbari, M. Shamsollahi, C. Jutten, and B. Coppa, ECG denoising using angular velocity as a state and an observation in an Extended Kalman Filter framework. *Proc. IEEE Conf. Eng. Med. Bio. Society* **1,** 2897-2900 (2012).

[33] R. Sameni, M. Shamsollahi, C. Jutten, and G. Clifford, A nonlinear Bayesian filtering framework for ECG denoising. *IEEE Trans. Biomed. Eng.* **54,** 2172-2185 (2007).

[34] R. Sameni, G. Clifford, C. Jutten, and M. Shamsollahi, Multi-channel ECG and noise modeling: Application to maternal and fetal ECG signals. *EURASIP J. Advances in Sig. Proc.* **43407,** 14pp (2007).

[35] F. Kazemzadeh, A. Wong, B. Behr, and A. Haijian, Depth Profilometry via Multiplexed Optical High-coherence Interferometry. *PLoS ONE* vol. 10, no. 3, e0121066 (2015).

[36] A. Wong, A. Mishra, K. Bizheva, and D.A. Clausi, General Bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery. *Optics Express* **18(8),** 8338-8352 (2010).

[37] S. Schwartz, C. Liu, A. Wong, D.A. Clausi, P. Fieguth, and K. Bizheva, Energy-guided learning approach to compressive FD-OCT. *Optics Express* **18(8),** 8338-8352 (2010).

[38] J. Glaister, A. Wong, and J. Glaister, "MSIM: Multistage illumination modeling of dermatological photographs for illumination-corrected skin lesion analysis," *IEEE Transactions on Biomedical Engineering* vol. 60, no. 7, pp. 1873-1883, 2013.

[39] Koh, D.M. and Padhani A.R. Diffusion-weighted MRI: a new functional clinical technique for tumour imaging. *Br. J. Radiol.* **79,** 633-635 (2006).

[40] Le Bihan, D., and Breton, E. Imagerie de diffusion in-vivo par resonance magnetique nucleaire. *C. R. Acad. Sci.* **301,** 1109-1112 (1985).

[41] M. J. Shafiee, S. Haider, A. Wong, D. Lui, A. Cameron, A. Modhafar, P. Fieguth and M. Haider, "Apparent Ultra-High b-value Diffusion-Weighted Image Reconstruction via Hidden Conditional Random Fields," IEEE Transactions on Medical Imaging, vol. 34, no. 5, pp. 1111-1124, 2015.

[42] A. Wong, X. Wang, and M. Gorbet, "Bayesian-based deconvolution fluorescence microscopy using dynamically updated nonparametric nonstationary expectation estimates," Sci. Rep. 1-13 (2015).

[43] A. Wong, F. Kazemzadeh, C. Jin, and X. Wang, "Bayesian-based aberration correction and numerical diffraction for improved lensfree on-chip microscopy of biological specimens ," Optics Letters vol. 10, no. 10, pp. 2233-2236, 2015.

[44] A. Wong, J. Glaister, A. Cameron, and M. Haider, Correlated Diffusion Imaging. *BMC Medical Imaging* **13,** 7pp (2013).

[45] A. Wong, F. Khalvati, and M. Haider, Dual-Stage Correlated Diffusion Imaging. *Proc. International Symposium on Biomedical Imaging (ISBI),* **75-78,** (2015).
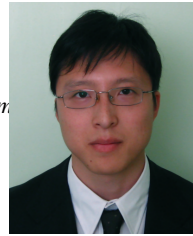
Alexander Wong (M' 05) received the B.A.Sc. degree in Computer Engineering from the University of Waterloo, Waterloo, ON, Canada, in 2005, the M.A.Sc. degree in Electrical and Computer Engineering from the University of Waterloo, Waterloo, ON, Canada, in 2007, and the Ph.D. degree in Systems Design Engineering from the University of Waterloo, ON, Canada, in 2010. He is currently the Canada Research Chair in Medical Imaging Systems, Co-director of the Vision and Image Processing Research Group, and an Assistant Professor in the Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada. He has published refereed journal and conference papers, as well as patents, in various fields such as computer vision, graphics, image processing, multimedia systems, and wireless communications. His research interests revolve around imaging, image processing, computer vision, pattern recognition, and cognitive radio networks, with a focus on integrative biomedical imaging systems design, probabilistic graphical models, biomedical and remote sensing image processing and analysis such as image registration, image denoising and reconstruction, image super-resolution, image segmentation, tracking, and image and video coding and transmission. Dr. Wong has received two Outstanding Performance Awards, an Engineering Research Excellence Award, an Early Researcher Award from the Ministry of Economic Development and Innovation, two Best Paper Awards by the Canadian Image Processing and Pattern Recognition Society (CIPPRS), a Distinguished Paper Award from Society for Information Display, and the Alumni Gold Medal.
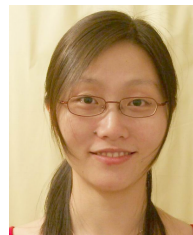
Xiao Yu Wang received the M.A.Sc. degree in Electrical Engineering from Concordia University, Montreal, Canada, in 2006, and the Ph.D. degree in Electrical and Computer Engineering from the University of Waterloo, ON, Canada, in 2011. She is currently an Adjunct Assistant Professor in the Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada. Her research interests include stochastic graphical learning and modeling for large-scale networks and data mining and visualization, affective computing, image processing, computer vision, signal processing, femtocell networking, network control theory, wideband spectrum sensing, and dynamic spectrum access. Her current focus is on efficient high-resolution, remote spatial biosignals measurements using video imaging for affective computing.