Signal Processing 131 (2017) 220–234

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Translation invariant multi-scale signal denoising based on goodness-of-fit tests

Naveed ur Rehman [a,*], Syed Zain Abbas [a], Anum Asif [a], Anum Javed [a], Khuram Naveed [a], Danilo P. Mandic [b]

[a] Department of Electrical Engineering, COMSATS Institute of Information Technology, Park Road, Chak Shahzad, Islamabad, Pakistan
[b] Department of Electrical and Electronic Engineering, Imperial College London, South Kensington, London SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 9 March 2016
Received in revised form
25 July 2016
Accepted 13 August 2016

Keywords:
Signal denoising
Discrete wavelet transform
Goodness of fit
Anderson Darling test statistic
Empirical distribution function

ABSTRACT

A novel signal denoising method based on discrete wavelet transform (DWT) and goodness of fit (GOF) statistical tests employing empirical distribution function (EDF) statistics is proposed. We cast the denoising problem into a hypothesis testing problem with a null hypothesis $\mathcal{H}_0$ corresponding to the presence of noise, and an alternative hypothesis $\mathcal{H}_1$ representing the presence of *only* desired signal in the samples being tested. The decision process involves GOF tests, employing statistics based on EDF, which is applied directly on multiple scales obtained from DWT. The resulting coefficients found to be belonging to noise are discarded while the remaining coefficients - corresponding to the desired signal - are retained. The cycle spinning approach is next employed on the denoised data to introduce translation invariance into the proposed method. The performance of the resulting method is evaluated against standard and modern wavelet shrinkage denoising methods through extensive repeated simulations performed on standard test signals. Simulation results on real world noisy images are also presented to demonstrate the effectiveness of the proposed method.

1. Introduction

Engineers and applied scientists routinely encounter data corrupted with unwanted noise which must be removed or reduced before further processing of the data. This process of removing noise, also known as denoising, has become a prerequisite step in many practical 1D and 2D (image) signal processing applications [1,2]. We present here a novel denoising method which operates by performing the goodness of fit (GOF) test on empirical discrete wavelet transform (DWT) coefficients; subsequently, the coefficients found to be corresponding to noise samples are removed. The denoising problem we attempt to address is given below.

Let the observed signal values be denoted by a vector $\mathbf{x}$, which is composed of a true signal $\mathbf{s}$ and an additive noise component $\boldsymbol{\eta}$

$$\mathbf{x} = \mathbf{s} + \sigma\boldsymbol{\eta}, \tag{1}$$

where $\boldsymbol{\eta}$ is considered to be a vector of independent random variables exhibiting Gaussian[1] distribution $\mathcal{N}(0, \sigma^2)$ with zero mean and arbitrary variance σ^2. The goal here is to estimate the true signal vector $\mathbf{s}$ from its noisy observations $\mathbf{x}$.

Early signal denoising techniques were based on the classical Weiner filter operating in the Fourier domain [3]. One major drawback of such methods was the inability of the Fourier transform to handle nonlinear and nonstationary data, which is often encountered in practice. Another class of denoising methods assumed the desired (denoised) signal to belong to a certain class of functions $\mathcal{L}$, and subsequently designing an estimator with a minimax risk. Based on that idea, several important denoising techniques [4–6] were developed; their scope, however, remained limited due to the difficulty in specifying a class of functions $\mathcal{L}$ for input data in hand, especially in practical scenarios.

These issues led to the development of denoising methods employing nonlinear operations, such as thresholding or shrinking, in the wavelet domain [7]. These methods have become extremely popular over the years due to their accuracy and ease of implementation and are being used as a de facto standard in applications involving signal denoising.

Wavelet shrinkage denoising methods operate by: i) taking the linear wavelet transform of a noisy input data; ii) estimating the threshold for each level from transform coefficients (required only for data-dependent thresholding); iii) performing non-linear thresholding in the wavelet domain; and iv) taking linear inverse wavelet transform.

Originally introduced by Donoho et al. [7,8], several variants of the wavelet shrinking denoising procedure can be generated by

* Corresponding author.

[1] While we focus entirely on Gaussian distributions in this paper, our proposed method is fully capable of handling arbitrary continuous noise distributions [52].

http://dx.doi.org/10.1016/j.sigpro.2016.08.019
0165-1684/

combining the different choices to implement the steps given above. Specifically, *VisuShrink* employing universal threshold for all levels, along with *RiskShrink* using universal minimax threshold, were proposed [8]. In [9], a local level-dependent thresholding method, *SUREShrink*, was proposed based on the Steins' Unbiased Risk Estimator (*SURE*). In [10], the process of nonlinear shrinkage of wavelet coefficients was modified to obtain nearly a minimax performance over a wide range of Triebel- and Besov-type smoothness constraints.

Lack of translation invariance of the wavelet basis, however, results in certain visual artifacts (especially near singularities) in the denoised signal obtained from the above mentioned standard wavelet denoising algorithms. Coifman and Donoho [11] proposed to average out that translation dependence in an attempt to reduce the artifacts, resulting in the so called second generation wavelet denoising methods. Their operation involves the following steps: i) shifting input data for a range of shift values; ii) performing denoising in each case; iii) un-shifting the resulting data; iv) averaging of the results.

Moreover, the threshold estimation procedures in standard wavelet denoising assume sparsity of wavelet coefficients due to the desired (deterministic) signal; this assumption, however, may not always be valid leading to erroneous threshold estimates and subsequently suboptimal denoising performances. To overcome this problem, an approach based on empirical Bayes method [12–14] in wavelet basis have been proposed which attempts to measure the extent of sparsity of wavelet coefficients by maximizing a marginal log likelihood function.

Further developments in wavelet based denoising led to a class of methods which exploited dependencies between wavelet coefficients, yielding improved performances [15–21]. These methods utilised the persistence property of empirical discrete wavelet coefficient across multiple scales. However, one problem encountered in such class of algorithms was over-smoothness of sharp features, such as bumps and discontinuities, within the output signal. Attempts to improve the denoising performance under such conditions gave rise to other approaches including total variation filtering [22], decompositions based on tracing sharp features in a signal [23], local wavelet based tree-structured estimators [21,24–26] and their adaptive version using dual tree complex wavelet transform (*DTCWT*) [27,28]. A very fast non-iterative algorithm for total variation filtering based denoising (*TVD*) has also been recently proposed [29].

A wavelet denoising approach employing shrinkage based on Bayes factors was proposed by Lavrik et al. [30]. The method is termed Bayesian Local False Discovery Rate (*BLFDR*), in which the underlying model on wavelet coefficients does not assume known variances. In the same paper, another algorithm, named Bayesian False Discovery Rate (*BaFDR*), is proposed which operates by ordering of posterior probabilities of hypotheses on empirical wavelets coefficients being null, in Bayesian testing framework of multiple hypotheses.

In [31], wavelet shrinkage (or thresholding) method was modified to be used in conjunction with Empirical Mode Decomposition (EMD) [32], leading to an iterative DWT-EMD thresholding algorithm (*TI-DWT-EMD*) with improved denoising performances for 1D signals. The method, however, did not utilise the statistical characteristics of white Gaussian noise and fractional Gaussian noise (fGn) for EMD, resulting in suboptimal results. To alleviate this problem, *TI-DWT-EMD* was modified in such that the desired modes were chosen based on the l_2 norm measure between the input and each mode [33]. EMD has also been employed separately for signal denoising and detrending problems [34]: the main idea is the partial reconstruction of relevant EMD components (modes), with those modes being chosen based on the empirically established statistical properties of EMD components for fractional Gaussian noise input. Despite the success of EMD in signal denoising, EMD lacks sound mathematical foundation and therefore its nonrecursive and mathematically well-founded variants such as variational mode decomposition [35] and synchrosqueezed wavelet transform [36–38] have been used for signal denoising to a good effect.

Sparse representations have been widely used in many signal processing algorithms recently, including signal denoising, restoration and reconstruction [39]. Early approaches, such as matching pursuit, assumed mutual independence among coefficients, which is suboptimal since inherent correlation among the coefficients is ignored. Moreover, for real world signals, the corresponding transform coefficients may not only exhibit sparsity but also clustering or grouping property; for instance, DWT coefficients generally have inter- and intra-scale grouping structures. A translation-invariant thresholding method exploiting the grouping properties of input signal/coefficient is presented in [40]. Model-based compressed sensing methods based on non-convex optimisation approaches have also emerged [41,42]. More recently, regularised methods employing non-greedy convex approaches have been successful in providing sparse estimates that retain a dependence structure [43].

Recently, multiscale denoising methods based on some similarity measures between the probability density functions (pdfs) of input and multiple scales of input data have emerged. One of those methods employs EMD and some similarity measure (both information theoretic and geometric measures had been explored in the paper) on pdfs to select those modes (scales) which correspond to noise [44]. In [33], the above idea was extended to a general case of fractional Gaussian noise (fGn). In both these methods, however, EMD had been employed which lacks solid mathematical foundation; secondly, both methods can be termed as 'global' methods, which attempt to find a 'noisy' mode through some pdf-based similarity measure. In most cases, noise is expected to be present in all modes and therefore a local approach is expected to be more suitable.

For that purpose, we present a novel method for denoising non-stationary data based on the DWT and GOF statistical tests using statistics based on empirical distribution function (EDF). The GOF statistical test is applied on short data segments of multiple input data scales, enabling a local framework for signal denoising. Moreover, we choose a linear DWT as a transform operator to obtain multiple scales, resulting in the same GOF based thresholds for all data scales.

The proposed method operates at multiple scales by identifying locally the DWT coefficients corresponding to noise and discarding those coefficients. The decision regarding the presence of noise samples in DWT scales is made using the GOF statistical tests, employing EDF statistics to quantify the similarity between observed coefficients with those expected under the model of DWT for white Gaussian noise (WGN) input. To achieve that, we formulate the signal denoising problem into a local hypothesis testing problem with the null hypothesis corresponding to the presence of noise and alternate hypothesis representing the case where only signal is present in the samples (wavelet coefficients) being tested. We further employ a cycle spinning approach [11] to average out the effects of translation dependence in the output signal, owing to the lack of translation invariance of orthogonal wavelet basis. Finally, we demonstrate the effectiveness of the proposed method by comparing it against the standard and state-of-the-art wavelet shrinkage denoising methods through extensive experiments for both 1D and 2D signals (images). It is emphasised that the proposed method differs from the standard wavelet based denoising methods in that it employs statistical GOF tests on multiple data scales to determine whether the DWT coefficients of a noisy input data correspond to noise or true