A novel image denoising method based on discrete wavelet transform (DWT) and goodness of fit (GOF) statistical tests employing empirical distribution function (EDF) statistics is proposed. We formulate the denoising problem into a hypothesis testing problem with a null hypothesis corresponding to the presence of noise, and alternate hypothesis representing the presence of only desired signal in the image samples being tested. The decision process involves GOF tests, employing statistics based on EDF, being applied directly on multiple image scales obtained from DWT. We evaluate the performance of the proposed method against the state of the art in wavelet image denoising through extensive experiments performed on standard images.

Index Terms— image denoising, wavelet transform, goodness of fit, empirical distribution function

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 signal and image processing applications. For 2D signals (images), the denoising problem can be formally specified as follows:

Let $x_{a,b}$ denote pixels of image $X$ at location $a, b = 1, \ldots, N$, which contains pixel values $s_{a,b}$ belonging to the true image $S$ along with additive noise $\eta_{a,b} \in \eta$

$$x_{a,b} = s_{a,b} + \eta_{a,b}$$

(1)

where $\eta$ is considered as an independent Gaussian noise $\mathcal{N}(0, \sigma^2)$, with zero mean and arbitrary variance $\sigma^2$. The goal here is to estimate $S$ from its noisy observation $X$.

Early signal denoising techniques were based on the classical Wiener filter operating in the Fourier domain. One major drawback of such methods was the inability of the Fourier transform to handle nonlinear and non-stationary data, often encountered in practice. These issues led to the development of multiscale methods employing nonlinear operations, such as thresholding or shrinkage, in the wavelet domain [1]. Such class of algorithms exploited the sparsity of desired signal values in the wavelet domain by specifying a threshold for signal detection at multiple scales. The coefficients found to be above the threshold were retained, while the remaining were discarded. While early methods based on this principle targeted 1D signals, their extensions for 2D signals or images soon followed.

Among 2D multiscale image denoising methods, the simplest is VIsuShrink [1] which employs universal threshold for all scales/bands, however, it over-smooths large sized images due to the dependence of its estimated threshold on image size. Better performances have been reported for denoising methods employing scale-adaptive threshold values, such as SureShrink [2], BayesShrink [3], and Surelet [4]. Another approach, employing empirical Bayesian estimation, exploited the sparsity and de-correlation properties of DWT for the purpose of image denoising [5]. Recently, a couple of techniques based on empirical Bayesian estimation have been proposed which employ 2D scale-mixing complex-valued wavelet transform [6] and hidden Bayesian network to model the prior probability of the original image [7].

Another class of multiscale image denoising methods exploit statistical dependencies among wavelet coefficients at different scales for estimating the threshold: BiShrink [8] models those dependencies based on non-Gaussian bivariate distributions. ProbShrink [9], on the other hand, estimates the probability that a given coefficient carries significant information by assuming a generalized Laplacian prior for noise free data. A recent method performs denoising through patch-based Wiener filters [10]. Similarly, BM3D [11] involves grouping of similar fragments of 2D transformed coefficients, arranging these fragments into 3D data arrays, followed by the attenuation of noise operation via spatial collaborative hard-thresholding. The computational complexity of the method is considerably large, however, owing to its complicated multi-step procedure.

We present here a novel image denoising method which operates by performing the goodness of fit (GOF) test locally on empirical discrete wavelet transform (DWT) coefficients of an image at multiple scales. For that cause, EDF statistics are used to quantify the similarity of the observed image coefficients with those expected from white Gaussian noise (WGN) input. Subsequently, the coefficients found to be sim-
ilar to noise samples are removed, while the remaining samples are retained. The established cycle spinning approach is also employed to render translation invariance property to the proposed approach [12]. Note that the proposed technique is fundamentally different from standard wavelet based image denoising as it involves performing hard thresholding on EDF statistics of wavelet coefficients rather than the coefficients themselves. The effectiveness of the resulting scheme has been shown by comparing it against the state of the art in multiscale image denoising via extensive experiments performed on standard test images.

2. RATIONALE OF WAVELET DENOISING

Standard wavelet denoising methods operate under the assumption that the wavelet coefficients due to the desired image $s_{a,b}$ are sparse in nature, leading to the distribution of total energy of $s$ in only a few coefficients in the wavelet domain, with relatively larger amplitudes. As a result, the majority of wavelet coefficients correspond to noise $\eta$ and have lower amplitudes. The rationale behind wavelet denoising is to exploit this sparsity of wavelet coefficients by removing (setting to zero) the coefficients belonging to noise. This is achieved by estimating and subsequently applying a nearly optimal threshold value, or for each scale separately.

The specification of $\tau_{AD}$ is falsely rejected; $P_{fa}$ denotes the probability of erroneously detecting a noise sample as one from desired signal, and must therefore be reduced as much as possible.

$$\text{Prob}\{\tau_{AD} > T | \mathcal{H}_0\} = P_{fa}$$

where $\mathcal{H}_0$ denotes the null hypothesis corresponding to the case of noise detection.

3. GOODNESS-OF-FIT TEST

Goodness-of-Fit (GOF) test indicates how well a specified model or distribution fits a given set of observations. Typically, GOF testing involves defining a measure of GOF, which is used to quantify the difference between the observed values and the values expected under the specified distribution. While many GOF measures are available presently with each having its unique properties [13], we will only discuss the widely used Anderson Darling (AD) [14] test statistics in this work.

Let us denote the empirical cumulative distribution function (ECDF) of input samples by $F(x)$ and the hypothesised cumulative distribution function (CDF) by $\mathcal{F}_r(x)$. The AD test statistic $\tau_{AD}$ is given by the following relation:

$$\tau_{AD} = \int_{-\infty}^{\infty} (F_r(x) - F(x))^2 \psi(F_r(x)) d(F_r(x))$$

where $\psi(F_r(x))$ is a non-negative weight function given by $(F_r(x)(1 - F_r(x)))^{-1}$, and is defined over the interval $0 \leq x < 1$. The relation for AD statistic, given in (2), can be rewritten in a more convenient form as:

$$\tau_{AD} = -W - S, \quad (3)$$

where

$$S = \sum_{i=1}^{W} \frac{(2i-1)\ln(F_r(x_i)) - \ln(F_r(x_{W+1-i}))}{W}$$

and $W$ denotes the length of window or segment (of data under observation) containing the ordered and normalized values $x_i$. The probability distribution of distances $\tau_{AD}$ are asymptotically specified for infinite window lengths, i.e. for $W \Rightarrow \infty$.

A threshold $T_\alpha$ could be computed for a given error probability of false alarm $P_{fa}$ under which the candidate distribution is falsely rejected; $P_{fa}$ denotes the probability of erroneously detecting a noise sample as one from desired signal, and must therefore be reduced as much as possible.

$$\text{Prob}\{\tau_{AD} > T | \mathcal{H}_0\} = P_{fa}$$

where $\mathcal{H}_0$ denotes the null hypothesis corresponding to the case of noise detection.

4. GOF BASED WAVELET DENOISING

We propose to address the denoising problem by performing the GOF tests locally on empirical wavelet coefficients of ‘noise + image’ mixture. For that purpose, a hypothesis testing framework is developed with the null $\mathcal{H}_0$ and alternative hypothesis $\mathcal{H}_1$ corresponding to the detection of noise only and true image pixels only cases, respectively.

$$\mathcal{H}_0 : \tau_{AD} < T_j$$
$$\mathcal{H}_1 : \tau_{AD} \geq T_j \quad (6)$$

where $T_j$ corresponds to the chosen threshold for scale $j$.

The specification of $\mathcal{H}_0$ and $\mathcal{H}_1$ as given in (6) requires obtaining discrete wavelet coefficients of noisy image at multiple scales. Next, the GOF test is performed locally on coefficients belonging to each scale, to determine if the observed coefficients belong to WGN distribution or the desired image pixels. The AD statistic provides robustness and flexibility compared to other GOF measures because it gives more weight to the tail of the distribution. Hence, the AD test statistic is employed within the GOF test to quantify the difference between the observed values and the values expected under reference noise distribution (Gaussian noise in our case). The resulting test statistic value $\tau$ is then compared against a threshold value $T_k$ - which is a function of the specified probability of false alarm $P_{fa}$. To determine if the coefficients belong to noise or desired image pixels. Those coefficients belonging to noise are discarded to yield the estimate of the true image. The steps involved in the proposed algorithm, we shall call it $GOF$-$DWT$, are discussed next:

4.1. Threshold Estimation

In the proposed $GOF$-$DWT$ method, for a given value of $P_{fa}$, the threshold $T_k$ must be computed for each scale level
4.2. DWT Decomposition

An input noisy image $X$ is then decomposed into its empirical wavelet coefficients as

$$W = W(X),$$

where $W$ is composed of the empirical wavelet coefficients $w^l(i, j)$ in accordance with Fig. 1. $k$ denotes the scale of decomposition and $l$ denotes the index of horizontal ($l = 1$), vertical ($l = 3$) or diagonal ($l = 2$) wavelet coefficients at a given scale.

4.3. Noise Variance Estimation

The GOF-DWT method requires the estimate of input noise level $\sigma$ for its operation. This is estimated by calculating the median of the diagonal DWT coefficients at the first level (corresponding to the finest scale) via the following relation

$$\hat{\sigma} = \frac{\text{median}(|\{w^1(i, j)\}_{i,j=1...,N}|)}{0.6745} \quad (10)$$

Subsequently, the wavelet coefficient vector $W$ is normalized by $\hat{\sigma}$

$$\tilde{W} = \frac{W}{\sigma} \quad (11)$$

to obtain the normalized wavelet coefficients $\tilde{w}^l(i, j)$ which are arranged in $\tilde{W}$ in accordance with Fig. 1.

4.4. GOF Thresholding using AD Statistic

In this step, the GOF test employing AD statistic is performed on each scale of noisy image to identify the empirical wavelet coefficients corresponding to WGN distribution, which are subsequently discarded. Let us denote the GOF thresholding operation by $G$, which is performed on wavelet coefficients $\tilde{w}^l(i, j)$ separately to yield denoised coefficients as follows

$$\tilde{W} = G(\tilde{W}, T_j). \quad (12)$$

where $\tilde{W}$ is composed of the denoised empirical wavelet coefficients $\tilde{w}^l(i, j)$ in accordance with Fig. 1.

The GOF operation $G$ is based on the local binary hypothesis testing problem modeled in (6): the decision regarding the presence of noise ($H_0$) or actual image pixels ($H_1$) is made for wavelet coefficients at each scale by applying the GOF based thresholding. Firstly, the GOF measure $\tau^{(i,j)}_{(k)}$ is computed for DWT coefficients $\tilde{w}^l(i, j)$ at indexes $\{i, j\}$ using (3), where the index set $\{i, j\}$ denotes the midpoint of an image segment of $W$ of size $5 \times 5$. Subsequently, for each index set and scale $k$, if $\tau^{(i,j)}_{(k)} < T_k$, then that coefficient in the $k$th scale is discarded (set to zero), else its original value is retained, yielding $\tilde{w}^l(i, j)$. Finally, the above two steps are repeated for all scales $k = 1 \ldots K$.

4.5. Translation Invariant Image Reconstruction

The denoised image $\hat{S}$ is reconstructed by applying the inverse DWT operation $W^{-1}$ on $\tilde{W}$ as follows

$$\hat{S} = W^{-1}(\tilde{W}), \quad (13)$$

where $\hat{S}$ is further scaled by the noise level estimate $\hat{\sigma}$, i.e. $\hat{S} = \hat{S} \times \hat{\sigma}$ to compensate for the normalisation performed earlier.

To compensate for the lack of translation invariance of the critically sampled DWT operation, 2D cycle spinning procedure is used, which applies a suitable inverse circulant shift
Fig. 2: Graphical illustration of the GOF-DWT method. The left hand side depicts the process of numerical estimation of threshold vs $P_{fa}$ relation. On the right hand column, GOF procedure is shown on one of the scales of the Barbara image. Note that the GOF operation is local, as it is performed on square window size of 5x5 pixels. The end result is the binary decision ($H_0$ or $H_1$) regarding the distribution of the window being considered.

operator to $\hat{S}$ and averages the subsequent results to obtain the final denoised image. Fig. 2 shows the block diagram of the GOF-DWT method, excluding the cycle spinning step. The proposed method after including the cycle spinning step will be referred to as TI-GOF-DWT throughout the paper.

5. EXPERIMENTAL RESULTS

In this section, we present results of extensive experiments performed using the proposed TI-GOF-DWT method and other state of the art denoising methods, operating on standard noisy input images. The images used in this case are the standard ‘Lena’, ‘Barbara’, ‘Aeroplane’ and ‘Medical’ images, which are shown in Fig. 3. The images were corrupted with the Gaussian noise of standard variations in the range of $\sigma = 10$ to $\sigma = 50$. The noisy input images were denoised using the TI-GOF-DWT method along with other wavelet based image denoising algorithms, including the SureShrink-TI SureTI [12], BiShrink BiShr [8], ProbShrink ProbShr [9], Surelet SureLet [4], and cSM [6]. Four decomposition levels of the Daubechies’ 8-tap filter, used as a mother wavelet, were used in all our experiments. The window size is chosen to be $5 \times 5$ pixels for local GOF operation in the TI-GOF-DWT; smaller window size results in poor estimates of CDF, whereas larger window size compromises the local analysis. The $P_{fa}$ was specified to be equal to 0.005. The parameters of the comparative denoising methods were all chosen to enable fair comparison with the TI-GOF-DWT. We performed repeated simulations with $K = 50$ iterations for each test image (and input variance) to obtain accurate quantitative results.

Table 1 shows the mean peak signal to noise ratios (PSNR) of the denoised images obtained from different denoising methods, for a range of input noise variances. The mean PSNR values were taken after performing $K = 50$ iterations at each instance. It is evident from Table 1 that the TI-GOF-DWT largely outperformed other methods on all input images for a range of input noise variances. Only Probshr was able to perform slightly better than the TI-GOF-DWT method for ‘Lena’ and ‘Aeroplane’ images at noise levels $\sigma = 10$ and $\sigma = 50$ respectively. In our experiments, the proposed method took 30 seconds on average to denoise a $256 \times 256$ noisy image.

To validate the above observation regarding the effectiveness of TI-GOF-DWT in denoising of images, we take the specific case of the ‘Medical’ image corrupted with noise of level $\sigma = 30$. In Fig. 4, in addition to the original and noisy images, the denoised images obtained from ProbShr, Surelet, cSM and TI-GOF-DWT are respectively shown in Fig. 4(c-f). Note that the TI-GOF-DWT (along with ProbShr) was able to capture the details of the original image better as compared to others, which is specifically evident in the upper part of the image.

Moreover, Fig. 5 shows the denoised ‘Lena’ images obtained from TI-GOF-DWT for different noise variances. Again, note that the proposed method was able to capture fine details of the ‘Lena’ image very effectively, even for high input noise standard deviation of $\sigma = 40$. 

![Fig. 3: Input images: (from left to right) Lena, Barbara, Aeroplane and Medical images used in our experiments.](Image)

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lena</td>
<td>512×512</td>
<td>512×512</td>
<td>512×512</td>
<td>512×512</td>
<td>512×512</td>
</tr>
<tr>
<td>SureTI</td>
<td>38.51</td>
<td>31.01</td>
<td>20.27</td>
<td>27.99</td>
<td>26.75</td>
</tr>
<tr>
<td>BiShr</td>
<td>34.29</td>
<td>31.10</td>
<td>20.09</td>
<td>27.72</td>
<td>26.62</td>
</tr>
<tr>
<td>ProbShr</td>
<td>34.44</td>
<td>31.48</td>
<td>20.29</td>
<td>27.02</td>
<td>25.32</td>
</tr>
<tr>
<td>Surelet</td>
<td>34.54</td>
<td>30.97</td>
<td>20.67</td>
<td>27.83</td>
<td>26.86</td>
</tr>
<tr>
<td>cSM</td>
<td>34.09</td>
<td>30.05</td>
<td>20.18</td>
<td>27.93</td>
<td>27.09</td>
</tr>
<tr>
<td>Prop.</td>
<td>34.04</td>
<td>31.42</td>
<td>20.67</td>
<td>27.83</td>
<td>27.29</td>
</tr>
<tr>
<td>Image</td>
<td>Lena</td>
<td>512×512</td>
<td>512×512</td>
<td>512×512</td>
<td>512×512</td>
</tr>
<tr>
<td>SureTI</td>
<td>38.50</td>
<td>30.89</td>
<td>20.87</td>
<td>27.49</td>
<td>26.17</td>
</tr>
<tr>
<td>BiShr</td>
<td>34.55</td>
<td>30.75</td>
<td>20.82</td>
<td>27.29</td>
<td>26.06</td>
</tr>
<tr>
<td>ProbShr</td>
<td>34.71</td>
<td>31.28</td>
<td>20.84</td>
<td>27.61</td>
<td>26.88</td>
</tr>
<tr>
<td>Surelet</td>
<td>34.71</td>
<td>30.85</td>
<td>20.86</td>
<td>27.45</td>
<td>28.47</td>
</tr>
<tr>
<td>cSM</td>
<td>34.04</td>
<td>30.49</td>
<td>20.90</td>
<td>27.51</td>
<td>26.91</td>
</tr>
<tr>
<td>Prop.</td>
<td>34.78</td>
<td>31.27</td>
<td>20.93</td>
<td>27.84</td>
<td>26.75</td>
</tr>
</tbody>
</table>
6. CONCLUSIONS

We have presented a wavelet-based image denoising method which employs the goodness of fit (GOF) statistical test to identify and subsequently suppress noise samples in discrete wavelet transform (DWT) coefficients of input image. Statistics based on empirical distribution function (EDF), Anderson Darling (AD) measure, has been employed in the GOF testing process. The presented approach, by design, can deal with arbitrary noise distributions though we have only focused on white Gaussian noise (WGN) in this work. The proposed approach has been shown to outperform the state of the art in wavelet-based image denoising methods via thorough experimental results.

7. REFERENCES