

MULTI-SCALE IMAGE DENOISING BASED ON GOODNESS OF FIT (GOF) TESTS

Naveed ur Rehman¹, Khuram Naveed¹, Shoaib Ehsan², Klaus McDonald-Maier²

¹Department of Electrical Engineering, COMSATS Institute of Information Technology, Park Road, Islamabad, Pakistan

²School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

ABSTRACT

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 \mathbf{X} at location $a, b = 1 \dots N$, which contains pixel values $s_{a,b}$ belonging to the true image \mathbf{S} along with additive noise $\eta_{a,b} \in \boldsymbol{\eta}$

$$x_{a,b} = s_{a,b} + \eta_{a,b} \quad (1)$$

where η is considered as an independent Gaussian noise $\mathcal{N}(0, \sigma^2)$, with zero mean and arbitrary variance σ^2 . The goal here is to estimate \mathbf{S} from its noisy observation \mathbf{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 η 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 either for all wavelet scales T (universal thresholding), or for each scale separately T_k (level-dependent thresholding).

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 $\mathcal{F}(x)$ and the hypothesised cumulative distribution function (CDF) by $\mathcal{F}_r(x)$. The AD test statistic τ_{AD} is given by the following relation:

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

where $\psi(\mathcal{F}_r(x))$ is a non-negative weight function given by $(\mathcal{F}_r(x)(1 - \mathcal{F}_r(x)))^{-1}$, 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

$$S = \sum_{i=1}^W \frac{(2i-1)}{W} (\ln(\mathcal{F}_r(x_i)) - \ln(\mathcal{F}_r(x_{W+1-i}))) \quad (4)$$

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 τ_{AD} are asymptotically specified for infinite window lengths, i.e. for $W \Rightarrow \text{inf}$.

A threshold T_α 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.

$$\text{Prob}\{\tau_{AD} > T | \mathcal{H}_0\} = P_{fa} \quad (5)$$

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.

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

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 τ 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

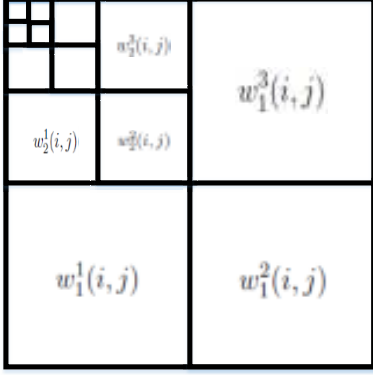


Fig. 1: Formation of the wavelet transformed empirical wavelet coefficients matrix W .

k . To find that relation, it is important to revisit the original relation of the AD statistic, as given in (2). It can be shown that the asymptotic distribution of τ_{AD} under \mathcal{H}_0 is

$$\mathcal{F}(\tau_{AD}|\mathcal{H}_0; x) = \frac{\sqrt{2\pi}}{\tau_{AD}} \sum_{i=1}^{\infty} a_i(4i+1) \exp\left(-\frac{(4i+1)^2\pi^2}{8\tau_{AD}}\right) \times \int_0^{\infty} \exp\left(\frac{\tau_{AD}}{8(W^2+1)} - \frac{(4i+1)(\pi w)^2}{8\tau_{AD}}\right) dW, \quad (7)$$

where $a_i = (-1)^i \Gamma(i+5)/(\Gamma(0.5)i!)$ and Γ denotes the gamma function [15]. For the AD statistic, using the definition of P_{fa} in (5), the relation between the threshold T and P_{fa} is given by:

$$P_{fa} = 1 - \mathcal{F}(\tau_{AD}|\mathcal{H}_0; x) \quad (8)$$

The speed of convergence of the limiting distribution in (7) is such that it suffices to use that distribution instead of using the actual distribution of τ_{AD} even for $W \geq 5$, while calculating threshold T as a function of P_{fa} [15]. Based on that, tables listing the values of thresholds T against the probability of false alarm P_{fa} are available [15]. Another approach is to obtain the relation between T_k and P_{fa} through numerical simulations on a large set of WGN data. This process, though not further discussed in detail in this paper due to space constraints, is graphically shown in Fig. 2.

4.2. DWT Decomposition

An input noisy image \mathbf{X} is then decomposed into its empirical wavelet coefficients as

$$\mathbf{W} = \mathcal{W}(\mathbf{X}), \quad (9)$$

where \mathbf{W} is composed of the empirical wavelet coefficients $w_k^l(i, j)$ 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 $\hat{\sigma}$ 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

$$\hat{\sigma} = \frac{\text{median}(|\{w_1^2(i, j)\}_{i,j=1,\dots,\frac{N}{2}}|)}{0.6745} \quad (10)$$

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

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

to obtain the normalized wavelet coefficients $\tilde{w}_k^l(i, j)$ which are arranged in $\tilde{\mathbf{W}}$ 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 \mathcal{G} , which is performed on wavelet coefficients $\tilde{w}_k^l(i, j)$ separately to yield denoised coefficients as follows

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

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

The GOF operation \mathcal{G} is based on the local binary hypothesis testing problem modeled in (6): the decision regarding the presence of noise (\mathcal{H}_0) or actual image pixels (\mathcal{H}_1) is made for wavelet coefficients at each scale by applying the GOF based thresholding. Firstly, the GOF measure $\tau_k^{(i,j)}$ is computed for DWT coefficients $\tilde{w}_k^l(i, j)$ at indexes $\{i, j\}$ using (3), where the index set $\{i, j\}$ denotes the mid point of an image segment of \mathbf{W} of size $W = 5 \times 5$. Subsequently, for each index set and scale k , if $\tau_k^{(i,j)} < T_k$, then that coefficient in the k th scale is discarded (set to zero), else its original value is retained, yielding $\hat{w}_k^l(i, j)$. Finally, the above two steps are repeated for all scales $k = 1 \dots K$.

4.5. Translation Invariant Image Reconstruction

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

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

where $\hat{\mathbf{S}}$ is further scaled by the noise level estimate $\hat{\sigma}$, i.e. $\hat{\mathbf{S}} = \hat{\mathbf{S}} \times \hat{\sigma}$ 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

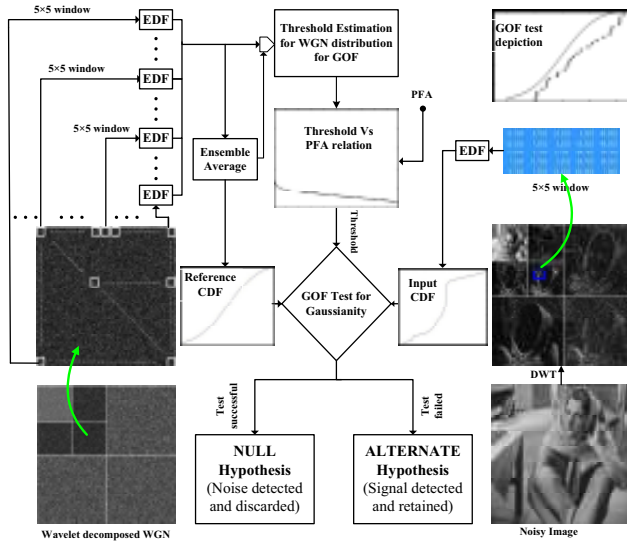


Fig. 2: Graphical illustration of the *GOF-DWT* method. The left hand side depicts the process of numerical estimation of threshold vs P_{fa} relation. On the right hand column, GOF procedure is shown on one of the scales of the Barbara image. Note that the GOF operation is local, as it is performed on square window size of 5×5 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 5×5 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 cho-



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

Table 1: Denoising results: Output PSNR values obtained against different input noise variances for input images.

σ	10	20	30	40	50	10	20	30	40	50
PSNR	28.13	22.11	18.59	16	14.15	28.13	22.11	18.59	16	14.15
Image	Lena (512 × 512)					Barbara (512 × 512)				
SureTI	34.31	31.09	29.27	27.99	26.75	32.71	28.56	26.34	24.93	23.91
Bishr	34.29	31.03	29.09	27.72	26.62	32.81	28.78	26.59	25.12	24.08
ProbShr	34.44	31.48	29.29	28	27.02	33.22	29.04	26.58	24.89	23.85
Surelet	34.34	30.97	29.07	27.81	26.86	32.46	28.17	25.97	24.65	23.83
cSM	34.09	30.95	29.18	27.97	27.05	32.53	28.57	26.46	25.15	24.27
Prop.	34.64	31.47	29.67	28.31	27.29	33.36	29.42	27.26	25.96	24.89
Image	Aeroplane (512 × 512)					Medical (256 × 256)				
SureTI	34.50	30.89	28.87	27.45	26.17	34.87	30.73	28.43	26.84	25.43
Bishr	34.35	30.75	28.67	27.20	26.06	34.24	30.19	27.96	26.42	25.27
ProbShr	34.71	31.25	29.14	27.67	26.66	35.11	30.99	28.66	27.10	25.90
Surelet	34.51	30.86	28.86	27.47	26.47	34.49	30.41	28.18	26.68	25.54
cSM	34	30.49	28.60	27.31	26.39	33	29.22	27.16	25.70	24.63
Prop.	34.78	31.23	29.23	27.84	26.75	39.07	35.15	33.05	31.55	30.37

sen 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 *Probshshr* 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 256×256 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 $\sigma = 40$.

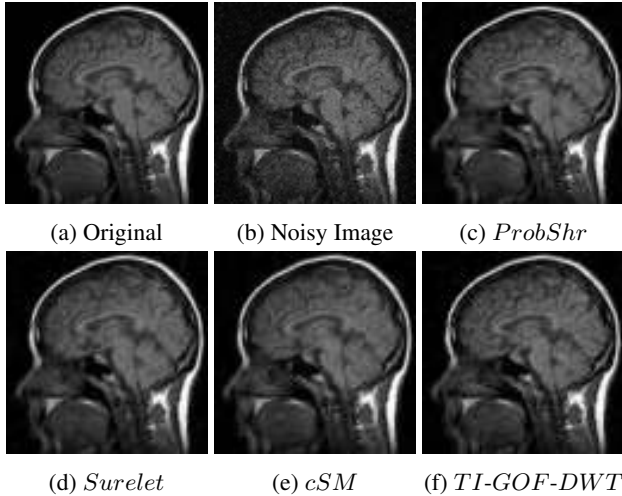


Fig. 4: Input, noisy ($\sigma = 30$) and denoised Medical images obtained from different methods.

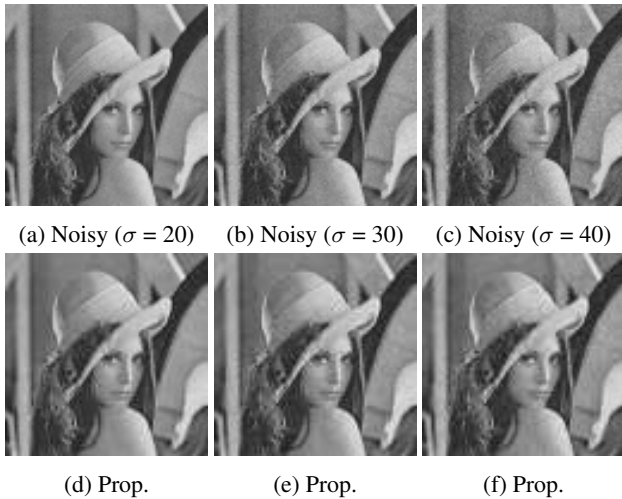


Fig. 5: Denoised Lena image obtained from the proposed *TI-GOF-DWT* method at different input noise variances.

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

- [1] D. L. Donoho, and I. M. Johnstone, "Minimax estimation via wavelet shrinkage", *Technical Report*, Department of Statistics, Stanford University, 1992.
- [2] D. L. Donoho, and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage", *Journal of the American Statistical Association*, vol. 90, pp. 1200-1224, 1995.
- [3] M. Vetterli, S. G. Chang, and B. Yu, "Adaptive wavelet thresholding for image denoising and compression", *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532-1546, 2000.
- [4] T. Blu, and F. Luisier. "The SURE-LET approach to image denoising," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2778-2786, 2007.
- [5] M. A. T Figueiredo and R. D. Nowak, "Wavelet based image estimation: An empirical Bayes approach using Jeffereys non-informative prior", *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1322-1331, 2001.
- [6] N. Remenyi, O. Nocolis, G. Nason, and B. Vidakovic, "Image Denoising with 2-D Scale-Mixing Complex Wavelet Transforms", *IEEE Transactions on Image Processing*, vol. 13, pp. 5165-5174, 2014.
- [7] J. Ho and W. L. Hwang, "Wavelet Bayesian network image denoising," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1277-1290, 2013.
- [8] L. Sendur, I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency", *IEEE Transactions on Signal Processing*, vol. 50, pp. 2744-2756, 2002.
- [9] A. Pizurica and W. Philips "Estimating probability of presence of a signal of interest in multiresolution single- and multiband image denoising", *IEEE Transactions on Image Processing*, vol. 15, pp. 654-665, 2006.
- [10] P. Chatterjee and P. Milanfar, "Patch-based near-optimal image denoising," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1635-1649, 2012.
- [11] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080-2095, 2007.
- [12] R. R. Coifman and D. L. Donoho, "Translation-invariant denoising", *Springer*, New York, 1995.
- [13] Y. Ingster, and I. A. Suslina, "Nonparametric Goodness-of-Fit Testing Under Gaussian Models", Springer Science and Business Media, 2003.
- [14] T. W. Anderson, and D. A. Darling, "A Test of Goodness-of-Fit", *Journal of the American Statistic Association*, vol. 49, pp. 765-769, 1954.
- [15] M. A. Stephens, "EDF statistics for goodness-of-fit and some comparisons", *Journal of the American Statistic Association*, vol. 69, pp. 730-737, 1974.