DENOISING IMAGES USING A NEW TYPE OF BISHRINK FILTER

ALEXANDRU ISAR[1], SORIN MOGA[2], DORINA ISAR[1]

Key words: MAP-filter, Double tree complex wavelet transform, Sensitivity.

This paper presents a new denoising method in the wavelet domain, which aims to reduce the noise, preserving the structural features (like the discontinuities) and textural information of the scene. In this paper we propose the association of the Double Tree Complex Wavelet Transform, (DT-CWT) with a Maximum A Posteriori (MAP) filter named bishrink. The corresponding denoising algorithm is simple and fast. Some simulation results and comparisons prove the performance of the new algorithm.

1. **INTRODUCTION**

The aim of a denoising algorithm is to reduce the noise level, while preserving the image features. In the wavelet domain, the noise is uniformly spread throughout the coefficients, while most of the image information is concentrated in the few largest ones (sparsity of the wavelet representation). Donoho and Johnstone propose a three steps denoising algorithm [1]:

1. the computation of a forward WT,
2. the filtering with a non-linear filter,
3. the computation of the corresponding inverse wavelet transform (IWT).

They use the Discrete Wavelet Transform (DWT) and the soft-thresholding filter. This filter puts to zero all the wavelet coefficients with the absolute value smaller then a threshold. This threshold is selected to minimize the min-max approximation error. The soft-thresholding filter was enhanced in [2–4]. A highly appealing particularity of the WTs is the inter-scale dependence. If at a given scale a coefficient is large, its correspondent at the next scale (having the same spatial coordinates) will be also large. In [2–4] the inter-scale dependencies are used to improve the denoising performance. The wavelet coefficients statistical models which exploit the dependence between coefficients give better results compared to the ones using an independent assumption [5, 6, 8, 9]. The denoising is performed in [5] and [6] with the aid of maximum a posteriori filters, (MAP). If we denote

[1] University "Politehnica" of Timişoara, E-mail: alexandru.isar@etc.upt.ro
[2] Telecom Bretagne, Brest, France

with w the wavelet coefficients of the noise-free image and with n the wavelet coefficients of the noise then it can be written, $y = w + n$. The MAP estimation of w, $\hat{w}$, realized using the observation y is given by the following equation:

$$\hat{w} = \text{argmax}\left\{log\left(p_n\left(y-w\right)p_w\left(w\right)\right)\right\}, \tag{1}$$

where p_x represents the probability density function (pdf) of x. For example if both w and n are Gaussian distributed,

$$p_w\left(w\right) = \frac{1}{\sqrt{2\pi}\ \sigma}\mathrm{e}^{-\frac{w^2}{2\sigma^2}};\ p_n\left(n\right) = \frac{1}{\sqrt{2\pi}\ \sigma_n}\mathrm{e}^{-\frac{n^2}{2\sigma_n^2}},$$

then the MAP filter becomes the very well known zero order Wiener filter. Its input-output relation (the solution of the MAP filter equation) is: $\hat{w} = \sigma^2 y/\left(\sigma^2+\sigma_n^2\right)$. To apply this relation, the two variances must be estimated. The precision of the estimation given in the last relation can be improved if a local estimation of the standard deviation of the noise-free image is considered. The model of naturally images is given by heavy tailed distributions, [5]. So, the utilization of zero-order Wiener filters in image denoising applications can not furnish the best performance. This drawback can be partially compensated by a better estimation of the local variance of the noise-free image, realized with the aid of two-stage denoising system [10–12]. The first stage treats the acquired image furnishing a pilot for the second stage. The acquired image is once again treated by the second stage but this time the information carried by the pilot is used. If n is Gaussian distributed and w has a Laplacian distribution then the MAP filter becomes an adaptive soft thresholding filter. In this case:

$$p_w\left(w\right) = \frac{1}{\sqrt{2}\sigma}\cdot\mathrm{e}^{-\frac{\sqrt{2}|w|}{\sigma}};\ p_n\left(n\right) = \frac{1}{\sqrt{2\pi}\sigma_n}\cdot\mathrm{e}^{-\frac{n^2}{2\sigma_n^2}}.$$

The MAP filter equation's solution becomes: $\hat{w} = \text{sgn}\left(y\right)\left(\left|y\right| - \sqrt{2}\frac{\sigma_n^2}{\sigma}\right)_+$,

where:

$$\left(X\right)_+ = \begin{cases} X & \text{for } X > 0, \\ 0 & \text{otherwise.} \end{cases}$$

This is a soft-thresholding filter with threshold value $\sqrt{2}\,\sigma_n^2/\sigma$. Once again the two standard deviations must be estimated. The zero order Wiener filter and the adaptive soft-thresholding filter are two examples of marginal MAP filters. If the

models of the clean image and of the noise are bivariate distributions then the MAP filter can take into account the inter-scale dependence of the wavelet coefficients. This is the case of the bishrink filter [5], the coefficient w_2 represents the parent of the coefficient w_1 (w_2 is the wavelet coefficient at the same position as w_1, but at the next coarser scale). Then: $y_k = w_k + n_k,\ k = 1, 2,$ and the vectors $\boldsymbol{w} = (w_1, w_2)$, $\boldsymbol{y} = (y_1, y_2)$ and $n = (n_1, n_2)$ can be constructed. With their aid it can be written: $\boldsymbol{y} = \boldsymbol{w} + \boldsymbol{n}$. The noise is assumed i.i.d. Gaussian,

$$p_n(n) = \frac{1}{2\pi\sigma_n^2} \mathrm{e}^{-\frac{n_1^2+n_2^2}{2\sigma_n^2}}.$$

The model of the noise-free image proposed in [5] is:

$$p_w(w) = \frac{3}{2\pi\sigma^2} \mathrm{e}^{-\frac{\sqrt{3}}{\sigma}\sqrt{w_1^2+w_2^2}}, \tag{2}$$

another heavy tailed distribution. The input-output relation of the bishrink filter is:

$$\hat{w}_1 = \frac{\left(\sqrt{y_1^2 + y_2^2} - \frac{\sqrt{3}\sigma_n^2}{\sigma}\right)_+}{\sqrt{y_1^2 + y_2^2}} y_1. \tag{3}$$

To estimate the noise variance from the noisy wavelet coefficients, a robust median estimator is used from the finest scale wavelet coefficients, [1]:

$$\hat{\sigma}_n^2 = \frac{\mathrm{median}\left(|y_i|\right)}{0.6745},\ y_i \in \text{subband HH}. \tag{4}$$

In [5] the marginal variance of the k-th coefficient is estimated using neighbouring coefficients in the region $N(k)$, a squared shaped window centred on this coefficient with size 7×7. To make this estimation one gets $\sigma_y^2 = \sigma^2 + \sigma_n^2$, where σ_y^2 represents the marginal variance of noisy observations y_1 and y_2. For the estimation of the marginal variance of noisy observations, in [5] is proposed the following relation:

$$\sigma_y^2 = \frac{1}{M} \sum_{y_i \in N(k)} y_i^2, \tag{5}$$

where M is the size of the neighborhood $N(k)$. Then σ can be estimated as:

$$\hat{\sigma} = \sqrt{\left(\hat{\sigma}_y^2 - \sigma_n^2\right)_+}. \tag{6}$$

In [6], a similar technique is used. The WT associated with the MAP filters described in [5] and [6] is no longer the DWT, because it has some drawbacks [7]: the lack of shift invariance, the lack of symmetry of the mother wavelets and the poor directional selectivity. These disadvantages can be diminished using a complex wavelet transform, like for example the DT-CWT [7]. In [8] is described a method for removing noise from digital images, based on a statistical model of the coefficients of overcomplete multiscale oriented basis. Neighbours of coefficients at adjacent positions (intra-scale dependence) and scales (inter-scale dependence) are modelled as the product of two independent random variables: a Gaussian vector and a hidden positive scalar multiplier. Under this model, named Gaussian scale mixture (GSM), the Bayesian least squares estimate, (BLS) of each coefficient reduces to a weighted average of the local linear estimate over all possible values of the hidden multiplier variable. In [9], three novel wavelet domain denoising methods for subband-adaptive, spatially-adaptive and multivalued image denoising are developed. The core of this approach is the estimation of the probability that a given coefficient contains a significant noise-free component, which is called signal of interest.

The aim of the present paper is to correct the comportment of the bishrink filter in the homogeneous regions of very noisy images. We propose a two-stage architecture. In the first stage a variant of the genuine denoising algorithm proposed in [5] is applied obtaining the first result. Computing the standard deviation of each pixel of the first result, the pilot image is obtained. At the basis of the construction of the second stage lies the idea of diversification. The structure of this paper is the following. In the second paragraph is presented a sensitivity analysis of the bishrink filter and some of its drawbacks are identified. Then a solution to reduce these drawbacks is proposed. In the third paragraph are given all the details of the proposed denoising algorithm. The fourth paragraph is dedicated to the presentation of the simulation results. The paper's conclusion is formulated in the fifth paragraph.

2. THE BISHRINK FILTER, ITS DRAWBACKS AND POSSIBLE ENHANCEMENTS

The estimator described by (2–6) is named bishrink filter and is applied in the field of the DT-CWT. The sensitivity of the bishrink filter with the estimation of the noise standard deviation $\hat{\sigma}_n$ can be computed with the relation:

$$S_{\hat{w}_1}^{\hat{\sigma}_n} = \frac{\mathrm{d}\,\hat{w}_1}{\mathrm{d}\,\hat{\sigma}_n}\frac{\hat{\sigma}_n}{\hat{w}_1}. \tag{7}$$

It is equivalent with:

$$S_{\hat{w}_1}^{\hat{\sigma}_n} = \begin{cases} \dfrac{-2\sqrt{3}\hat{\sigma}_n^2}{\hat{\sigma}\sqrt{y_1^2+y_2^2}-\sqrt{3}\hat{\sigma}_n^2}, & \text{if } \sqrt{y_1^2+y_2^2} > \dfrac{\sqrt{3}\hat{\sigma}_n^2}{\hat{\sigma}}, \\ 0, & \text{in rest.} \end{cases} \tag{8}$$

The absolute value of this sensitivity is an increasing function of $\hat{\sigma}_n$. Another very important parameter of the bishrink filter is the local estimation of the marginal variance of the noise-free image, $\hat{\sigma}$. The sensitivity of the estimation $\hat{w}_1$ with $\hat{\sigma}$ is given by:

$$S_{\hat{w}_1}^{\hat{\sigma}} = \begin{cases} \dfrac{\sqrt{3}\hat{\sigma}_n^2}{\hat{\sigma}\sqrt{y_1^2+y_2^2}-\sqrt{3}\hat{\sigma}_n^2}, & \text{if } \sqrt{y_1^2+y_2^2} > \dfrac{\sqrt{3}\hat{\sigma}_n^2}{\hat{\sigma}}; \\ 0, & \text{in rest.} \end{cases} \tag{9}$$

This is a decreasing function of $\hat{\sigma}$. So, the bishrink filter treats very well the edges, the estimation of the textured regions must be corrected and the worst treatment corresponds to the homogeneous regions. The sensitivity $S_{\hat{w}_1}^{\hat{\sigma}}$ increases with the increasing of $\hat{\sigma}_n$. In consequence the most difficult regime of the bishrink filter corresponds to the treatment of homogeneous regions of very noisy images. To reduce this drawback we propose to use the diversification principle [10]. We consider three diversification mechanisms: the utilization of two different mother wavelets and the utilization of two different variants of bishrink filter, named: adaptive bishrink filter with global estimation of local variance and mixed bishrink filter. Six estimates, called in the following partial results, are obtained. The final estimate is obtained by the fusion of the six partial results. Using the six masks generated at the end of the first stage, we identify in each partial result the six corresponding classes. Each one contains only the pixels with the coordinates specified by the corresponding mask. The procedure proposed prevents the oversmoothing using a different number of partial results in regions with different homogeneity degree.

3. THE PROPOSED SOLUTION

The architecture of the proposed denoising kernel is presented in Fig. 1. The first stage of the algorithm contains the blocks: DT-CWT_A, F2, IDTCWT_A and Segm. The second one contains the other blocks. Two types of WT, DT-CWT_A and DT-CWT_B are computed obtaining the wavelet coefficients w_A and w_B. Three variants of bishrink filter, F_1-the genuine one, F_2-the adaptive bishrink filter with global estimation of the local variance and F_3-the mixed bishrink filter, are applied in the field of each DT-CWT. Six estimates of wavelet coefficients $\hat{w}_{1A}, \hat{w}_{2A}, \hat{w}_{3A}, \hat{w}_{1F}, \hat{w}_{2F}$ and $\hat{w}_{3F}$ are obtained. For each one the inverse WT, IDT-CWT, is computed, obtaining six partial results, $\hat{s}_{1A}, \hat{s}_{2A}, \hat{s}_{3A}, \hat{s}_{1F}, \hat{s}_{2F}$ and $\hat{s}_{3F}$. The image $\hat{s}_{2A}$ is segmented in six classes following the values of the local variances of its pixels obtaining the pilot image. Registering the coordinates of the pixels belonging at each of these classes the six masks are generated by the block Segm. With the aid of these masks the six classes of each partial result are identified. Using the class selectors CS_1-CS_6, each partial result is treated in a different manner. These systems pick up the pixels of their input images with the coordinates belonging to the correspondent mask, generating each class of the partial results. The first class of the final estimate $\hat{s}_1$ is identical with the first

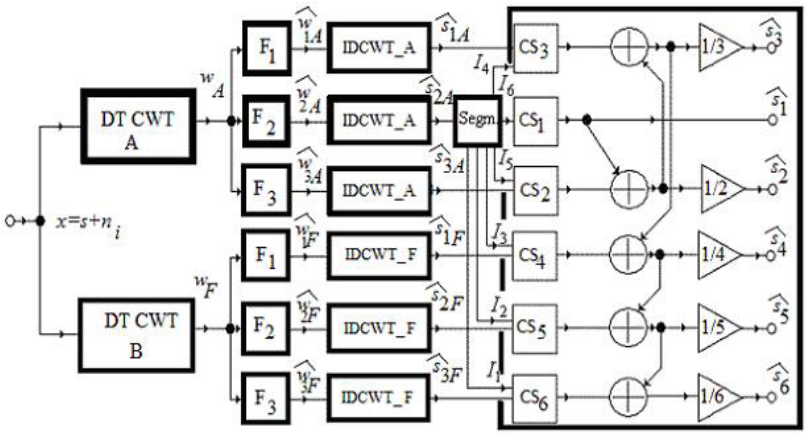

Fig. 1 – The architecture of the proposed additive noise denoising kernel.

class of the image $\hat{s}_{2A}$. The second class of the final result, $\hat{s}_2$, is obtained averaging the pixels belonging to the second classes of the partial results $\hat{s}_{2A}$ and $\hat{s}_{3A}$ and so on.

3.1. **THE DIVERSIFICATION MECHANISMS**

The first diversification mechanism refers to the construction of the DT-CWT. Since an image usually consists of several regions of different smoothness, the sparsity of its representation in a single wavelet domain is limited. This naturally motivates using multiple wavelet transforms to denoise [11]. There are two kinds of filters used for the computation of the DT-CWT: for the first decomposition level and for the other levels [7]. The first diversification mechanism is realized by the selection of two types of filters for the first level. The first one is selected from the (9, 7)-tap Antonini filters pair and the second one corresponds to the pair of Farras nearly symmetric filters for orthogonal 2-channel perfect reconstruction filter bank [12]. The other diversification mechanisms refer to the construction of bishrink filter variants and represent original contributions. The filter F_1 is a genuine bishrink filter. The filter F_2 is a bishrink filter with global estimation of the local variance [10]. It was constructed for the reason presented in the following. The estimation in (5) is not very precise. It is based on the correct assumption that y_1 and y_2 are modelled as zero mean random variables. But their restrictions to the finite neighbourhood $N(k)$ are not necessary zero mean random variables. So, it is better to estimate first the means in the neighbourhood:

$$\hat{\mu}_y = (1/M). \sum_{y_i \in N(k)} y_i, \tag{10}$$

and then the variances:

$$\hat{\sigma}_y^2 = (1/M). \sum_{y_i \in N(k)} (y_i - \hat{\mu}_y)^2. \tag{11}$$

Finally, the relation (5) can be applied.

In the case of the bishrink filter with global estimation of the local variance, the detail wavelet coefficients produced by the first tree of the DT-CWT computation block are indexed with “re” and the detail wavelet coefficients produced by the other tree are indexed with “im”. Applying in order the relations (10), (11), (5), (4) and (6) for the two trees implementing each of the DT-CWTs, the local parameters: ${}_{\text{re}}\hat{\mu}_y$, ${}_{\text{re}}\hat{\sigma}_y^2$, ${}_{\text{re}}\hat{\sigma}$, ${}_{\text{im}}\hat{\mu}_y$, ${}_{\text{im}}\hat{\sigma}_y^2$ and ${}_{\text{im}}\hat{\sigma}$ are computed in each neighbourhood $N(k)$. Then the global estimation of the marginal standard deviation can be done: $\hat{\sigma} = ({}_{\text{re}}\hat{\sigma} + {}_{\text{im}}\hat{\sigma})/2$.

The filter F_3 is the mixed bishrink filter, proposed in [10].

3.2. **THE CLASSIFICATION**

The image $\hat{s}_{2A}$ is segmented in classes which elements have a value of the local variance $\hat{\sigma}_{2A}$, belonging to one of six possible intervals: $I_p = \left(\alpha_p \hat{\sigma}_{2\,A\max}, \alpha_{p+1} \hat{\sigma}_{2\,A\max}\right)_{1\le p\le 6}$, where $\alpha_1 = 0$ and $\alpha_7 = 1$. The class selector CS_p in Fig. 1 selects the class associated to the interval I_{7-p}. For uniform regions $\hat{\sigma}$ is proportional with $\hat{\sigma}_{2A}$. Preliminary extensive tests proved that the six estimates are classified from better to poor in the following sequence: $\hat{s}_{2A}$, $\hat{s}_{3A}$, $\hat{s}_{1A}$, $\hat{s}_{1F}$, $\hat{s}_{2F}$ and $\hat{s}_{3F}$, from the peak signal to noise ratio (PSNR) point of view. These tests also suggest the following values for the bounds of the intervals I_p: $\alpha_2 = 0.025, \alpha_3 = 0.05, \alpha_4 = 0.075, \alpha_5 = 0.1$ and $\alpha_6 = 0.25$.

4. **SIMULATION RESULTS**

We compare, in Table 1, the proposed denoising method to other effective systems in the literature, namely the systems proposed in [2, 11, 5, 6, 8, 9]. The comparison was done using for images: Lena, Boat and Barbara, all having the same size 512 × 512 pixels. Let s and $\hat{s}$ denote the clean and the denoised images. The root mean square (rms) of the approximation error is given by: $\varepsilon = \sqrt{(1/N)\sum_q \left(s_q - \hat{s}_q\right)^2}$, where N is the number of pixels. The PSNR in dB is expressed by: $\mathrm{PSNR} = 20\log_{10}\left(\frac{255}{\varepsilon}\right)$. The algorithm described in [6] has two implementations, one using local estimations ("with l.e") and the other using only global estimations ("Not l.e"). The cases not analyzed in the references already indicated were marked in Table 1 as "–". Analyzing this table some observations can be made. For all the test images and all noise levels, with only one exception (Barbara, $\sigma_n = 100$) the better results are obtained using the BLS-GSM algorithm. The proposed algorithm can be found on the second place. The distance between the performances of these algorithms is of approximately 0.25 dB in the case of the image Lena and of approximately 0.3 dB in the case of the image Barbara, for the levels of noise usually considered in literature, $5 < \sigma_n < 35$. For the very noisy cases, $35 < \sigma_n$, this distance decreases and for the image Barbara when $\sigma_n = 100$, the output PSNR obtained using the proposed denoising method is higher than the output PSNR obtained using the BLS-GSM algorithm.

Table 1

The PSNR values of denoised results for different test images and noise levels (σ_n) of (A) – noisy, (B) – [2], (C) – [11], (D) – [5], (E) – [6], (F) – [8], (G) – [9] and (H) – proposed system

	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Lena					Not l.e.			
σ_n=10	28.18	-	-	35.34	34.75	**35.61**	35.24	35.36
σ_n=15	24.65	-	-	33.67	33.03	**33.90**	33.46	33.68
σ_n=20	22.14	-	-	32.40	31.87	**32.66**	32.20	32.43
σ_n=25	20.17	-	29.4	31.40	30.89	**31.69**	31.21	31.49
σ_n=30	18.62	-	-	30.54	30.18	-	30.33	30.83
σ_n=50	14.15	-	-	28.27	-	**28.61**	-	28.56
σ_n=100	8.13	-	-	25.18	-	**25.64**	-	25.47
Boat					with l.e.			
σ_n=10	28.15	32.90	-	33.10	33.09	**33.58**	33.25	33.33
σ_n=15	24.62	30.85	-	31.36	31.44	**31.70**	31.32	31.45
σ_n=20	22.10	29.47	-	30.08	30.19	**30.38**	29.93	30.14
σ_n=25	20.17	28.44	28.6	29.06	29.21	**29.37**	28.89	29.12
σ_n=30	18.58	27.63	-	28.31	28.51	-	28.04	28.38
σ_n=50	24.74	25.50	-	26.01	-	**26.38**	-	26.12
σ_n=100	22.44	22.97	-	23.31	-	**23.75**	-	23.45
Barbara								
σ_n=10	28.13	32.18	-	33.53	-	**34.03**	33.46	33.78
σ_n=15	24.61	29.66	-	31.31	-	**31.86**	31.19	31.57
σ_n=20	22.11	27.98	-	29.80	-	**30.32**	29.53	30.03
σ_n=25	20.17	26.76	26.7	28.61	-	**29.13**	28.23	28.88
σ_n=30	18.62	25.83	-	27.65	-	-	27.17	**27.93**
σ_n=50	14.15	23.70	-	25.40	-	**25.48**	-	25.45
σ_n=100	8.13	21.76	-	22.54	-	22.61	-	22.63

5. **CONCLUSION**

This paper presents an efficient image denoising algorithm based on a two-stage architecture. The result obtained at the output of the first stage is segmented on the basis of its local variance into six regions having different homogeneity degrees. Each region is diversified using a diversification order inverse proportional with its homogeneity degree, to reduce the sensitivity $S_{\hat{w}_1}^{\hat{\sigma}}$. Then the corresponding region is synthesized by averaging. The proposed algorithm improves the treatment of homogeneous zones for very noisy images. The diversity enhancement technique proposed is more general. It can be applied with minor modifications to other denoising architectures. We presented our simulation results and compared with published results. The comparisons suggest that the results obtained are competitive with the best results reported in the literature.

Received on 11 November 2008

REFERENCES

1. D. L. Donoho, I. M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Statist. Assoc., **90**, *432*, pp. 1200–1224, 1995.
2. F. Luisier, T. Blu, M. Unser, *A New SURE Approach to Image Denoising: Interscale Orthonormal Wavelet Thresholding*, IEEE Transactions on Image Processing, **16**, *3*, pp. 593-606, 2007.
3. Z. Cai, T. H. Cheng, C. Lu, K. R. Subramanian, *Efficient wavelet-based image denoising algorithm*, Electronics Letters, **37**, *11*, pp. 683-685, 2001.
4. S. Chang, B. Yu, M. Vetterli, *Adaptive wavelet thresholding for image denoising and compression*, IEEE Transactions on Image Processing, **9**, *9*, pp. 1522-1531, 2000.
5. L. Sendur, I. W. Selesnick, *Bivariate shrinkage with local variance estimation*, IEEE Signal Processing Letters, **9**, *12*, pp. 438-441, 2002.
6. A. Achim, E. E. Kuruoglu, *Image Denoising Using Bivariate α-stable Distributions in the Complex Wavelet Domain*, IEEE Signal Processing Letters, **12**, *1*, pp. 17-20, 2005.
7. N. Kingsbury, *Complex Wavelets for Shift Invariant Analysis and Filtering of Signals*, Appl. and Comput. Harmonic Analysis, **10**, pp. 234-253, 2001.
8. J. Portilla, V. Strela, M. J. Wainwright, E. P. Simoncelli, *Image denoising using scale mixtures of gaussians in the wavelet domain*, IEEE Trans. Image Process., **12**, *11*, pp. 1338–1351, 2003.
9. A. Pizurica and W. Philips, *Estimating the probability of the presence of a signal of interest in multiresolution single-and multiband image denoising*, IEEE Trans. Image Process., **15**, *3*, pp. 645–665, 2006.
10. A. Isar, D. Isar and A. Quinquis, *Multi-Scale MAP Denoising of SAR and SAS Images*, Sea Technology, **48**, *2*, pp. 46-48, 2007.
11. H. Choi, R. G. Baraniuk, *Multiple Wavelet Basis Image Denoising Using Besov Ball Projections*, IEEE Signal Processing Letters, **11**, *9*, pp. 717-720, 2004.
12. A. Abdelnour, I. W. Selesnick, *Nearly symmetric orthogonal wavelet basis*, Proc. IEEE Int. Conf., Acoust., Speech, Signal Processing (ICASSP), 2001.