Image Denoising Using a Bishrink Filter with Reduced Sensitivity

Alexandru Isar [1], Sorin Moga[2], Dorina Isar[1]
[1] Communications Dept., Electronics and Telecommunications Faculty, "Politehnica" University, Timisoara, Romania
alexandru.isar@etc.upt.ro, dorina.isar@etc.upt.ro
[2] GET/ ENST Bretagne - TAMCIC/CNRS UMR 2872, Brest, France
sorin.moga@enst-bretagne.fr

***Abstract*—The performance of image denoising algorithms using the Double Tree Complex Wavelet Transform, DT CWT, followed by a local adaptive bishrink filter can be improved by reducing the sensitivity of that filter with the local marginal variance of the wavelet coefficients. In this paper is proposed a solution for the sensitivity reduction based on enhanced diversity. First the advantages and disadvantages of a state-of-the-art denoising solution, based on the association DT CWT - bishrink filter are highlighted. Second a blind noise-suppression method correcting the disadvantages of the bishrink filter, performing a non-linear operation on the data is obtained. Finally, some simulation examples prove the performances of the proposed denoising method.**

I. INTRODUCTION

During acquisition and transmission, images are often corrupted by additive noise that can be modeled as Gaussian most of the time. The aim of an image-denoising algorithm is then to reduce the noise level, while preserving the image features. Such a system must realize a great noise reduction in the homogeneous regions and the preservation of the details of the scene in the other regions. There is a great diversity of estimators used like denoising systems. A possible classification criterion for these systems takes into account the theory that is found at the basis of each one. In this respect, there are two categories of denoising systems, those based on wavelet theory and the others. In fact, David Donoho introduced the word denoising in association with the wavelet theory, [1]. From the first category, taking into account their performance, we must mention the denoising systems proposed in [2] and [3]. The denoising system proposed in [2] is based on the shape-adaptive DCT (SA-DCT) transform **that can be computed on a support of arbitrary shape**. The SA-DCT is used in conjunction with the Anisotropic Local Polynomial Approximation (LPA) - Intersection of Confidence Intervals (ICI) technique, **which defines the shape of the transform support in a pointwise adaptive manner.** Since supports corresponding to different points are in general overlapping, **the local estimates are averaged together using adaptive weights that depend on the region's statistics.** The denoising system proposed in [3] is a maximum *a posteriori* (MAP) filter that acts in the spatial domain. It makes a different treatment of regions with different homogeneity degree. **These regions can be treated independent with the same MAP filter choosing between different prior models.**

The multi-resolution analysis performed by the wavelet transform, (WT) has been shown to be a powerful tool to achieve good denoising. 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), [4-7]. The corresponding denoising methods have three steps, [1]: 1) The computation of the forward WT, 2) the filtering of the wavelet coefficients, 3) the computation of the inverse wavelet transform of the result obtained, (IWT). Numerous WTs can be used to operate these treatments. The first one was the Discrete Wavelet Transform, DWT, [1]. It has three main disadvantages, [8]: lack of shift invariance, lack of symmetry of the mother wavelets and poor directional selectivity. These disadvantages can be diminished using a complex wavelet transform. In the following, the Dual Tree Complex Wavelet Transform, DT CWT, [8, 9], will be used. This is a redundant WT, with a redundancy of 4. All the WTs have two parameters: the mother wavelets, MW and the primary resolution, PR, (number of iterations). The importance of their selection is highlighted in [10]. Another appealing particularity of those transforms, becoming from their multiresolution capability, is **the interscale dependency of the wavelet coefficients.** Numerous non-linear filter types can be used in the WT domain. A possible classification is based on the nature of the useful component of the image to be processed. Basically, there are two categories of filters: those built supposing that the useful component of the input image is deterministic and those based on the hypothesis that this component is random. To the first category belong the hard-thresholding filter, [1], the soft-thresholding filter, [1,11], that minimizes the Min-Max estimation error and the Efficient SURE-Based Inter-scales Pointwise Thresholding filter [11], that minimizes the Mean Square Error, (MSE). Filters obtained by minimizing a Bayesian risk, typically under a quadratic cost function (a delta cost function (maximum a posteriori-MAP estimation [4,**5**,7,12]) or the minimum mean squared error - MMSE estimation [6, 13]) belong to the second category. The construction of MAP filters supposes the existence of two statistical models, for the useful component of the input image and for its noise component. The MAP estimation of w, realized using the observation $y=w+n$, (where n represents the WT of the noise and w the WT of the useful component of the input image) is given by the following MAP filter equation:

$$\hat{w}(y) = \underset{w}{arg\,max}\{ln(f_n(y\text{-}w) f_w(w))\}$$

where f_x represents the probability density function (pdf) of x. Generally, the noise component is supposed Gaussian distributed. For the useful component there are many models. We have proved in [14] that this distribution changes from scale to scale. For the first iterations of the WT it is a heavy tailed distribution and with the increasing of iterations number it converges to a Gaussian. There are two solutions to deal with this mobility. The first one supposes to use a fixed simple model, risking an increasing of imprecision across the scales. This way, there is a chance to obtain a closed form input-output relation for the MAP filter. This is the case of the bishrink filter [5]. An explicit input-output relation has two advantages: it simplifies the implementation of the filter and it permits the analysis of its sensitivities. The second solution supposes to use a generalized model, defining a family of distributions and the identification of the best fitting element of this family for the distribution of the wavelet coefficients at a given scale. For example in [4] is used the family of Pearson's distributions, in [7] the family of SαS distributions and in [12] is used the model of Gauss–Markov random field. The use of such generalized model makes the treatment more precise but implies implicit solution for the MAP filter equation, it can be solved only numerically, demanding more time and memory resources, and the sensitivities of the filter obtained cannot be evaluated. If the pdfs f_w and f_n do not take into account the interscale dependency of the wavelet coefficients than the MAP filter obtained is called marginal. This paper proposes a new denoising method for images based on the association of the DT CWT with a filter bank composed by different variants of bishrink filters. With the aid of those filters the diversity is enhanced in the wavelets domain. This gain in diversity allows us to locally correct some distortions produced by the association DT CWT – bishrink filter. In fact, we propose a statistical segmentation of the result obtained applying the association DT CWT – bishrink, in regions with different homogeneity degree. Each such region is treated with a different element of the filter bank. Finally the results obtained this way are averaged together. The second section presents the architecture of the proposed denoising system. The aim of the third section is the presentation of some simulation results.

II. THE PROPOSED DENOISING METHOD

We consider the denoising of an image s corrupted by additive white Gaussian noise, AWGN, with variance σ_n^2. In order to exploit the interscale dependency of the wavelet coefficients, let w_{2k} represent the parent of w_{1k}. The parent is located at the same geometrical coordinates like the child, but at the successive scale. The problem is formulated in wavelet domain as $y_{1k} = w_{1k} + n_{1k}$ and $y_{2k} = w_{2k} + n_{2k}$. We can write:

$$\mathbf{y}_k = \mathbf{w}_k + \mathbf{n}_k. \tag{1}$$

where $\mathbf{w}_k = (w_{1k}, w_{2k})$, $\mathbf{y}_k = (y_{1k}, y_{2k})$ and $\mathbf{n}_k = (n_{1k}, n_{2k})$. In [5] is proposed a Laplace bivariate model for the wavelet coefficients of the useful component of the input image and a bivariate Gaussian model for the wavelet coefficients of the noise. The MAP estimator of w_1 derived using these models is:

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

Here $(g)_+$ is defined as:

$$(g)_+ = \begin{cases} 0, & if\ g < 0 \\ g, & \text{otherwise} \end{cases}. \tag{3}$$

This estimator, named bishrink filter, requires the prior knowledge of the noise variance σ_n^2, and the marginal variance σ^2, of the useful component of the input image, for each wavelet coefficient. To estimate the noise variance σ_n^2 from the noisy wavelet coefficients, a robust median estimator is used from the finest scale wavelet coefficients:

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

In [5], the marginal variance of the k'th coefficient is estimated using neighboring coefficients in the region $N(k)$, a squared shaped window centered at the k'th coefficient with window size 7x7. To make this estimation one gets $\sigma_y^2 = \sigma^2 + \sigma_n^2$, where σ_y^2 is the marginal variance of noisy observations, y_1 or y_2. For the estimation of σ_y^2, is proposed the following relation:

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

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

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

The estimator described by (2), (4) and (6) is named local adaptive bishrink filter. One of its most important parameters is the marginal variance, σ. The sensitivity of the estimation $\widehat{w}_1$ with $\hat{\sigma}$ is defined by:

$$S_{\widehat{w}_1}^{\hat{\sigma}} = \frac{d\widehat{w}_1}{d\hat{\sigma}} \cdot \frac{\hat{\sigma}}{\widehat{w}_1}. \tag{7}$$

For the coefficients satisfying:

$$\sqrt{y_1^2 + y_2^2} > \frac{\sqrt{3}\hat{\sigma}_n^2}{\hat{\sigma}}, \tag{8}$$

this sensitivity becomes:

$$S_{\widehat{w}_1}^{\hat{\sigma}} = \frac{\sqrt{3}\hat{\sigma}_n^2}{\hat{\sigma}\sqrt{y_1^2 + y_2^2} - \sqrt{3}\hat{\sigma}_n^2}. \tag{9}$$

It is inverse proportional with the local degree of homogeneity measured by the value of $\hat{\sigma}$.

The goal of this paper is to reduce the distortion of zones with different degree of homogeneity of the image produced by a denoising system based on the DT CWT and the local adaptive bishrink filter. The solution proposed is the enhancement of the estimation diversity. Two types of DTCWT are computed. For each wavelet coefficient, three variants of bishrink filter are applied, obtaining six different estimations $\hat{w}_1$. Averaging these values, a better estimation is obtained. For the wavelet coefficients with higher $\hat{\sigma}$ a reduced number of variants is applied. This procedure is equivalent with the use of *p* different denoising systems in the region corresponding to values of $\hat{\sigma}$ belonging to the interval I_{7-p} (defined latter), and the fusion of their results. The architecture of the proposed denoising system is presented in figure 1. Six estimates of the wavelet coefficients $\hat{w}_{1A}$, $\hat{w}_{2A}$, $\hat{w}_{3A}$, $\hat{w}_{1F}$, $\hat{w}_{2F}$ and $\hat{w}_{3F}$ are produced. For each one the IDTCWT is computed, obtaining six estimates, $\hat{s}_{1A}$, $\hat{s}_{2A}$, $\hat{s}_{3A}$, $\hat{s}_{1F}$, $\hat{s}_{2F}$ and $\hat{s}_{3F}$. The image $\hat{s}_{2A}$ is segmented in six classes following the values of the local standard deviations of its pixels. Using the class selectors $CS_1 - CS_6$ each image $\hat{s}_{1A}$ - $\hat{s}_{3F}$ is treated in a different manner. The segmentation block, Segm, creates the map of each class (the list of pixels positions belonging to that class). The corresponding class selectors, CS_p, use these maps. They pick up the pixels of their input images with the positions belonging to the corresponding map, generating the class of those images. CS_1 has only one input and generates the first class of the image $\hat{s}_{2A}$. CS_2 has two inputs and generates the second class of the images $\hat{s}_{2A}$ and $\hat{s}_{3A}$ and so on. The first class of the final estimate $\hat{s}_1$ is identical with the first 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 images $\hat{s}_{2A}$ and $\hat{s}_{3A}$ and so on. For the last class of the final result, $\hat{s}_6$, containing uniform zones, all the pixels belonging to the sixth class of the estimates, $\hat{s}_{1A}$, $\hat{s}_{2A}$, $\hat{s}_{3A}$, $\hat{s}_{1F}$, $\hat{s}_{2F}$ and $\hat{s}_{3F}$ are averaged. The filter F_1 is of bishrink type. The construction of the filters F_2 and F_3 correspond to two diversification principles: the estimation of the local standard deviations of the wavelet coefficients and the pdf of those coefficients. There are two kinds of filters used for the computation of the DT CWT: for the first level and for the other levels, [8]. The first diversification in figure 1 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 (for DT CWT A) and the second one (for DT CWT F) corresponds to the pair of Farras nearly symmetric filters for orthogonal 2-channel perfect reconstruction filter bank, [15]. The estimation of σ in (6) is not very precise for two reasons. First, it is based on the correct assumption that y_1 and y_2 are modeled as zero mean random variables. But their restrictions to the finite neighborhood *N (k)* are not necessary zero mean random variables. So, is better to

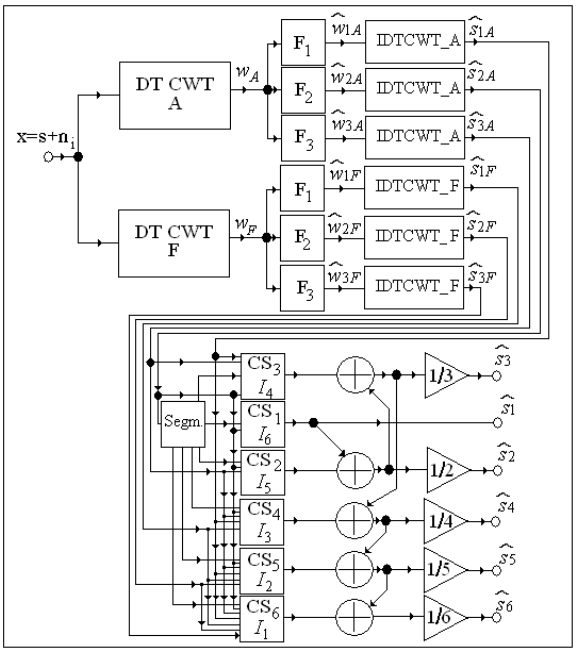


Figure 1. The architecture of the proposed denoising system.

estimate first the means in the neighborhood:

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

and then the variances:

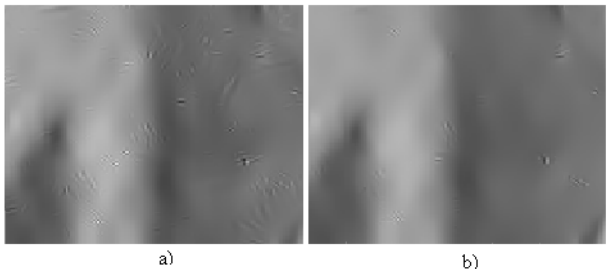
$$\hat{\sigma}_y^2 = \frac{1}{M} \sum_{y_i \in N(k)} \left(y_i - \hat{\mu}_y \right)^2, \tag{11}$$

Finally, the relation (6) can be applied. The second reason of imprecision is the fact that relation (6) refers only to one of the two trees of the DT CWT. In the following the detail wavelet coefficients produced by this three will be indexed with *re*. The detail wavelet coefficients produced by the other tree will be indexed with *im*. Applying, in order, the relations (10), (11) and (6), for the two trees implementing each of the DT CWT, the local parameters ${}_{re}\hat{\mu}_y$, ${}_{re}\hat{\sigma}_y^2$, ${}_{re}\hat{\sigma}$, ${}_{im}\hat{\mu}_y$, ${}_{im}\hat{\sigma}_y^2$ and ${}_{im}\hat{\sigma}$ are computed in each neighborhood *N(k)*. Then the global estimation of the marginal standard deviation obtained by averaging ${}_{re}\hat{\sigma}$ and ${}_{im}\hat{\sigma}$ is done. The filter F2 is a variant of the bishrink filter based on this estimation of the marginal standard deviation. In [14] is proposed a new variant of bishrink filter, named mixed bishrink filter, that acts for the first three iterations of each DWT like a bishrink filter, for the forth iteration like a local adaptive Wiener filter and for the fifth iteration of each DWT (the last one) like a hard thresholding filter with the threshold equal with $3\hat{\sigma}_n$. The filter F3 in figure 1 is a mixed bishrink filter. The image $\hat{s}_{2A}$ is segmented in classes whose elements have a value of the local variance, $\hat{\sigma}_{2A}$, belonging to one of six possible intervals $I_p = \left(\alpha_p \hat{\sigma}_{2A\max}, \alpha_{p+1} \hat{\sigma}_{2A\max} \right)_{p=\overline{1,6}}$, where: $\alpha_1 = 0, \quad and \quad \alpha_7 = 1.$ The class selector CS_p in figure 1, selects the class associated to the interval I_{7-p}. Preliminary tests proved that the six

estimates are classified from better to poor in the form: $\hat{s}_{2A}$, $\hat{s}_{3A}$, $\hat{s}_{1A}$, $\hat{s}_{1F}$, $\hat{s}_{2F}$ and $\hat{s}_{3F}$, from the pick signal to noise ratio, PSNR, point of view. These tests also suggest the following values for the bounds of the intervals I_p: $\alpha_2 = 0.025$, $\alpha_3 = 0.05$, $\alpha_4 = 0.075$, $\alpha_5 = 0.1$, $\alpha_6 = 0.25$. The first class of the final result contains some pixels of the image $\hat{s}_{2A}$. The second class of the final result is obtained by averaging the second classes of two partial results, $\hat{s}_{2A}$ and $\hat{s}_{3A}$ and so on. The sixth class of the final result is obtained by averaging the sixth classes of all the partial results, $\hat{s}_{2A}$, $\hat{s}_{3A}$, $\hat{s}_{1A}$, $\hat{s}_{1F}$, $\hat{s}_{2F}$ and $\hat{s}_{3F}$.

III. SIMULATION RESULTS

An example, highlighting the better performance of the new denoising algorithm in the uniform zones, is given in figure 2 for the image Lena. The original image was perturbed with an AWGN with $\sigma_n = 100$. A region obtained cropping the image $\hat{s}_{2A}$ is illustrated in figure 2 a). The same region was also extracted from the image $\hat{s}$ and is illustrated in figure 2 b). Analyzing the two pictures in figure 2 it can be observed that some very localized distortions, present in picture a), was corrected in picture b). We also compared the proposed algorithm to other effective systems in the literature, namely the local adaptive bishrink filter in [5], the denoising system based on the steerable pyramid proposed in [13] and the denoising processor introduced in [7]. The comparison was done using two images: Boats and Barbara, having the same size, 512x512 pixels and the results are presented in Table 1.



a) b)

Figure 2. The proposed denoising system b) corrects some distortions introduced by the bishrink filter a).

TABLE I. THE PSNR VALUES OF DENOISED IMAGES FOR DIFFERENT TEST IMAGES AND NOISE LEVELS (σ_n) OF (A) NOISY, (B) DENOISING SYSTEM IN [13], (C) DENOISING PROCESSOR IN [7], (D) LOCAL ADAPTIVE BISHRINK FILTER IN [5] AND (E) PROPOSED ALGORITHM.

σ_n	A	B	C	D	E
Boats					
10	28.16	-	33.09	33.10	**33.33**
15	24.65	-	31.44	31.36	**31.45**
20	22.14	-	**30.19**	30.08	30.14
25	20.17	-	**29.21**	29.06	29.12
30	18,62	-	**28.51**	28.31	28.38
Barbara					
10	28.16	33.45	-	33.35	**33.78**
15	24.65	31.22	-	31.31	**31.57**
20	22.14	29.71	-	29.80	**30.03**
25	20.17	28.57	-	28.61	**28.88**
30	18,62	-	-	27.65	**27.93**

IV. CONCLUSION

The results presented illustrate the effectiveness of the proposed algorithm. The comparison made suggests the new denoising results are competitive with the best wavelet-based results reported in literature. So, the less precise statistical model can be locally corrected obtaining a fast denoising algorithm. One of our future research directions is the formalization of the heuristic choices in this paper.

ACKNOWLEDGMENT

The research with the results reported in this paper was realized in the framework of the Programme de recherche franco-roumain Brancusi with the title: Débruitage des images SONAR en utilisant la théorie des ondelettes.

REFERENCES

[1] D. L. Donoho, I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage", Biometrika, 81(3) : 425-455, 1994.

[2] Foi, A., V. Katkovnik, and K. Egiazarian, "Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color Images", (in review) IEEE Trans. Image Process., April 2006. (preprint).

[3] M. Walessa, M. Datcu, "Model-Based Despeckling and Information Extraction from SAR Images", IEEE Transactions on Geoscience and Remote Sensing, vol. 38, no. 5, September 2000, 2258- 2269.

[4] Samuel Foucher, Gozie Bertin Benie, Jean-Marc Boucher, "Multiscale MAP Filtering of SAR images", IEEE Transactions on Image Processing, vol. 10, no.1, January 2001, 49-60.

[5] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency", IEEE Trans. on Signal Processing, 50(11): 2744-2756, November 2002.

[6] A. Pizurica, W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single and multiband image denoising", IEEE Transactions on Image Process. 15(3), 2006, 654-665.

[7] A. Achim and E. E. Kuruoglu, "Image Denoising Using Bivariate α-Stable Distributions in the Complex Wavelet Domain", IEEE Signal Processing Letters, 12(1):17-20, January 2005.

[8] N. Kingsbury, "Complex Wavelets for Shift Invariant Analysis and Filtering of Signals", Applied and Comp. Harm. Anal. 10, 2001, 234-253.

[9] N G Kingsbury, "A Dual-Tree Complex Wavelet Transform with improved orthogonality and symmetry properties", Proc. IEEE Conf. on Image Processing, Vancouver, September 11-13, 2000, paper 1429.

[10] G.P. Nason, "Choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage", Statistics and Computing, 12, 2002, 219-227.

[11] F. Luisier, T. Blu, M. Unser, "A New SURE Approach to Image Denoising: Inter-Scale Orthonormal Wavelet Thresholding," IEEE Transactions on Image Processing, in press.

[12] D. Gleich, M. Datcu, "Gauss–Markov Model for Wavelet-Based SAR Image Despeckling", IEEE Sig. Proc. Let., vol. 13, no. 6, June 2006, 365-368.

[13] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image Denoising using Scale Mixtures of Gaussians in the Wavelet Domain," IEEE Transactions on Image Processing, vol. 12, no. 11, November 2003.

[14] A. Isar, S. Moga, D. Isar, "A New Method for Denoising SONAR Images", Proceedings of IEEE International Symposium SCS'05, Iasi, Romania, July 14-15, 2005, 469-472.

[15] A. Abdelnour and I. W. Selesnick, "Nearly symmetric orthogonal wavelet bases", Proc. IEEE Int. Conf. Ac., Sp., Sig. Proc. (ICASSP), May 2001.