Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 49 (63), Fascicola 1-2, 2004

Denoising Over-Sampled Signals

André Quinquis[1], Alexandru Isar[2], IEEE Members and Dorina Isar[2]

Abstract - This paper presents a new denoising method for over-sampled constant within intervals signals corrupted by additive noise. The novelty of this paper is a special MAP filter, called composed bishrink. A complete statistical analysis of this filter is reported. Some simulations are presented. The results obtained are compared with the results of other denoising methods and with other state-of-the art filtering techniques.
Keywords: wavelets, denoising, soft thresholding, bishrink.

I. INTRODUCTION

In recent years, the techniques that use multiscale and local transform-based algorithms have become popular in noise filtering applications. In particular the use of non-linear filters in the DCT domain was studied, [1]. In this paper, we consider local transform based denoising. We propose such an algorithm, based on the discrete wavelet transform, (DWT). Section II deals with the local DWT-based denoising.
In section III, a statistical analysis of the new denoising method is presented. The use of local filters in the DWT domain is described in the following section. In section V, numerical simulation results are presented and discussed. The last section is dedicated to some concluding remarks.

II. LOCAL DWT-BASED DENOISING

The following model of the observed signal corrupted by additive noise is considered in this paper:

$$x[k] = s[k] + n[k] \qquad (1)$$

where s and n represent the useful part and the noise. The problem is to estimate s starting from x. The noise is usually considered to be an uncorrelated with s, stationary random process, with a null mean and a variance $\sigma_n{}^2$. To estimate the signal s, Donoho, [2], proposed the following method:
1. The Discrete Wavelet Transform (DWT) of the signal x is computed. The result is the signal $y_i = y + n_y$. The noise n_y converges asymptotically to a Gaussian white one, with the same variance, [3].
2. A non-linear filtering is applied in the wavelet domain:
where t is a threshold. This system is called soft

$$y_0[k] = \begin{cases} \operatorname{sgn}\{y_i[k]\}\left(|y_i[k]| - t\right), & |y_i[k]| > t \\ 0, & if \quad not \end{cases} \qquad (2)$$

thresholding filter. Because the noise n_y is Gaussian, if $t>3\sigma_n$, the probability $P(n_y>t)$ is very little (the rule of 3 sigmas). So the noise is quasi entirely suppressed. This is the reason why the signal y_0 is a denoised version of the signal y_i. This is a non-linear adaptive filter whose statistic analysis was presented in [4]. The adaptability is due to the selection of the threshold value in function of the noise power.
3. Taking the inverse DWT (IDWT) of the signal y_0, the denoised version of the signal s, x_0, is obtained.
The principal disadvantage of the already described denoising method is due to the fact that it is based only on the estimation of the noise variance (the useful part of the input signal is ignored) and on hypotheses confirmed only asymptotically. This is the reason why in the following another denoising strategy, based on the use of a Maximum a Posteriori, MAP, filter, will be described.

III. A STATISTICAL ANALYSIS OF THE DWT

The probability density function, (pdf), of the wavelet coefficients at the m'th scale, (after m iterations), ${}_xD_m^k$ (k being equal with 1 for detail coefficients and with 2 for approximation coefficients) is given by the following relation:

$$f_{{}_xD_m^k}(a) = \mathop{\ast}_{r_1=1}^{N(k)} \mathop{\ast}_{q_2=1}^{M_0} \cdots \mathop{\ast}_{q_m=1}^{M_0} f_d\left(k, r_1, q_2 ..., q_m, a\right) \qquad (3)$$

where:

2) Communications Department, "Politehnica" University Timisoara, e-mail: isar@hermes.ee.utt.ro

$$f_d\left(k,r_1,q_2...,q_m,a\right)= \\ =G\left(k,r_1,q_2...,q_m\right)\cdot \\ \cdot f_x\left(G\left(k,r_1,q_2...,q_m\right)a\right) \quad (4)$$

and:

$$G\left(k,r_1,q_2...,q_m\right)= \\ =\frac{1}{F\left(k,r_1\right)\prod_{l=2}^{m} m_o\left[q_l\right]} \quad (5)$$

where:

$$F\left(k,r_1\right)=\begin{cases} m_0\left[r_1\right] & k=2 \\ m_1\left[r_1\right] & k=1 \end{cases} \quad (6)$$

M_0 represents the length of the impulse response m_0 (of the low-pass filter used in the computation of the DWT) and M_1 represents the length of m_1 (the high-pass filter used in the computation of the DWT) and the number of the convolutions in the first group from the relation (3) is given by:

$$N\left(k\right)=\begin{cases} M_0, & k=2 \\ M_1, & k=1 \end{cases} \quad (7)$$

In conformity with (3), the pdf of the wavelet coefficients is a sequence of convolutions. Hence, the random variable representing the wavelet coefficients can be written like a sum of independent random variables. So, the central limit theorem can be applied. This is the reason why the pdf of the wavelet coefficients tends asymptotically to a Gaussian, when the number of convolutions in (3) tends to infinity. This number depends on the mother wavelets used and on the number of iterations of the DWT. For mother wavelets with a long support, this number increases very fast. The slower convergence is obtained for the Haar mother wavelets, which has the shorter support. This is the case analyzed in this paper. The filter used in the DWT domain must be constructed having in mind that after a number of iterations the distribution of the wavelet coefficients can be considered Gaussian. The problem is to establish this number, Nu_1. Another problem is to find the wavelet coefficients distribution law for the first iterations (before to reach the Gaussian law). The correlation of the wavelet coefficients ${}_xD_m^k$ is given by:

$$\Gamma_{{}_xD_m^k}\left[n_1,p_1\right]=E\left\{{}_xD_m^k\left[n_1\right]\left({}_xD_m^k\left[p_1\right]\right)^*\right\}= \\ \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{p_2=-\infty}^{\infty}\gamma_x\left(2^{-m}\left(w+2p_2\pi\right)\right)\cdot \\ \cdot e^{-jw\left(n_1-p_1\right)}\cdot\left|\Im\left\{\psi^k\right\}\left(w+2p_2\pi\right)\right|^2 \mathrm{d}w \quad (8)$$

Because the signals *s* and *n* are not correlated it can be written:

$$\Gamma_{{}_xD_m^k}=\Gamma_{{}_sD_m^k}+\Gamma_{{}_nD_m^k} \quad (9)$$

If the input noise is a zero mean white Gaussian, the correlation of its wavelet coefficients becomes:

$$\Gamma_{{}_nD_m^k}\left[n_1\right]=\sigma_n^2\cdot\delta\left[n_1\right] \quad (10)$$

At any scale the noise in the wavelets domain is also white, having the same variance. So, a single estimation of its variance, for example using the detail coefficients obtained after the first iteration, is sufficient. For any type of input noise, when *m* tends to infinity, the relation (8) becomes:

$$\Gamma_{{}_nD_\infty^k}\left[n_1,p_1\right]=\gamma_x\left(0\right)\cdot\ \delta\ \left[n_1-p_1\right] \quad (11)$$

So, asymptotically, the noise in the wavelets domain becomes white. Unfortunately this is also only an asymptotic result. Combining this result with the result obtained after the pdf analysis, it can be observed that after a given number of iterations, Nu_2, the noise in the wavelet domain is white and Gaussian. The mean of the wavelet coefficients is:

$$E\left\{{}_xD_m^k\left[n_1,p_1\right]\right\}=\begin{cases} 0, & k=1 \\ 2^{\frac{m}{2}}\cdot\mu_x, & k=2 \end{cases} \quad (12)$$

In practice the number of iterations of the DWT is high. The length of the approximation coefficients sequence obtained after the last iteration is small. This is the reason why this sequence is not filtered in practice. The variance of the wavelet coefficients is given by:

$$\sigma^2_{{}_xD_m^k}=E\left\{\left|{}_xD_m^k\right|^2\right\}= \\ =\frac{1}{2\pi}\int_{-\infty}^{\infty}\gamma_x(2^m u)\cdot\left|\Im\left\{\psi^k\right\}(u)\right|^2 du \quad (13)$$

The correlation of the DWT of the useful part of the input signal, *s*, is given by:

$$\Gamma_{{}_sD_m^k}\left[n_1\right]=2^m\cdot\Gamma_s\left[2^m n_1\right] \quad (14)$$

its mean by:

$$E\left\{{}_sD_m^k\left[n_1\right]\right\}=\begin{cases} 0, & k=1 \\ 2^{\frac{m}{2}}\cdot\mu_s, & k=2 \end{cases} \quad (15)$$

and its variance by:

$$\sigma^2_{{}_sD_m^k}=2^m\cdot\sigma_s^2 \quad (16)$$

So, the variances of the detail wavelet coefficients sequences of the useful component of the input signal increases when the iteration index increases.

IV. MAP FILTERS EXPLOITING THE INTERSCALE DEPENDENCY OF THE DETAIL COEFFICIENTS

In conformity with (8), there is an important correlation between a wavelet coefficient at a given scale and the same coefficient situated in the same position at the next scale (named the parent of the considered coefficient). This correlation can be exploited to construct adaptive filters acting at a given scale and using for the estimation of their parameters information obtained at the next scale, [5]. Using the parent and child wavelet coefficient of the input signal it is possible to estimate the child coefficients of the DWT of the useful part of the input signal, with the aid of a bishrink filter, [5]. Let ${}^1y_i$ be the considered detail coefficient and ${}^2y_i$ its parent. The statistical parameters of the child coefficients can be determined using their parent coefficients and the neighbor child coefficients, located in a window with a length of 3, centered on the current child coefficient. It can be written:

$$\mathbf{y}_i = \mathbf{y} + \mathbf{n}_y \qquad (17)$$

where:

$$\mathbf{y}_i = \left({}^1y_i, {}^2y_i\right); \mathbf{y} = \left({}^1y, {}^2y\right); \mathbf{n}_y = \left({}^1n_y, {}^2n_y\right) \qquad (18)$$

The MAP estimation of $\mathbf{y}$, realized using the observation $\mathbf{y}_i$, is given by:

$$\hat{\mathbf{y}}(\mathbf{y}_i) = \underset{\mathbf{y}}{\arg\max}\left\{\ln\left(f_{\mathbf{n}_y}(\mathbf{y}_i - \mathbf{y}) \cdot f_{\mathbf{y}}(\mathbf{y})\right)\right\} \qquad (19)$$

In the following, we will consider that the DWT of the noise is distributed following a zero mean Gaussian:

$$f_{\mathbf{n}_y}(\mathbf{n}_y) = \frac{1}{\sqrt{2\pi}\sigma_n} \cdot e^{-\frac{\left({}^1n_y\right)^2 + \left({}^2n_y\right)^2}{2\sigma_n^2}} \qquad (20)$$

Concerning the model of the DWT of the useful component, in the case of the composed bishrink filter, for the first Nu_2 iterations, a Laplace distribution will be considered (like in the case of the bishrink filter, [5]):

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{\sqrt{3}}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{\sqrt{3}}{\sigma}\cdot\sqrt{\left({}^1y\right)^2 + \left({}^2y\right)^2}} \qquad (21)$$

and for the other iterations, a Gaussian distribution will be considered:

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{1}{\sqrt{2\pi\,{}^1\sigma\,{}^2\sigma}} \cdot e^{-\frac{\left({}^1y\right)^2 + \left({}^2y\right)^2}{2\,{}^1\sigma\,{}^2\sigma}} \qquad (22)$$

(like in the case of the Wiener filter, [6]). For the models in (20) and (21) the solution of the maximization problem in (19) is:

$$\widehat{{}^1y} = \frac{\left(\sqrt{\left({}^1y_i\right)^2 + \left({}^2y_i\right)^2} - \frac{\sqrt{3}\,\widehat{\sigma_n}^2}{\hat{\sigma}}\right)_+}{\sqrt{\left({}^1y_i\right)^2 + \left({}^2y_i\right)^2}} \cdot {}^1y_i \qquad (23)$$

where:

$$\hat{\sigma}^2 = \widehat{{}^1\sigma} \cdot \widehat{{}^2\sigma} \qquad (24)$$

and:

$$(g)_+ = \begin{cases} g, & g > 0 \\ 0, & if \quad not \end{cases} \qquad (25)$$

and for the models in (20) and (22) the solution of the maximization problem in (19) is:

$$\widehat{{}^1y} = \frac{\widehat{{}^1\sigma} \cdot \widehat{{}^2\sigma}}{\widehat{{}^1\sigma} \cdot \widehat{{}^2\sigma} + \widehat{\sigma_n}^2} \cdot {}^1y_i \qquad (26)$$

So, the input-output relations of the composed bishrink filter are (23) and (26). The noise variance is estimated using the details obtained after the first iteration and the variances $\widehat{{}^1\sigma}$ and $\widehat{{}^2\sigma}$ are estimated in moving windows centered on the current child and parent coefficients. First the means are estimated in each window and second the variances. But, applying the relation (16), a different estimation of the local variance of the child coefficients can be obtained:

$$\widehat{{}^1\sigma_d} = \frac{\widehat{{}^2\sigma}}{\sqrt{2}} \qquad (27)$$

To profit of these two estimations of the local variances, obtained at two successive scales, it can be written:

$$\widehat{\widehat{{}^1\sigma}} = \frac{\widehat{{}^1\sigma} + \frac{\widehat{{}^2\sigma}}{\sqrt{2}}}{2} \qquad (28)$$

This estimation will be used in (24) and (26), substituting $\widehat{{}^1\sigma}$, for the input-output relations of the composed bishrink filter.

V. SIMULATION RESULTS

A useful input signal constant within intervals was considered. This is a data sequence, specific for the communication in the base-band. This sequence has a number of 16384 symbols, each having *b*=128 samples. A portion of this signal is represented in figure 1. Taking into account the waveform of this signal, the Haar mother wavelets must be used. The DWT was computed on blocks, each having a length of 4096 samples. The maximal number of iterations (equal with 12) was used for the computation of each DWT. For the implementation of the composed bishrink, a value of Nu_2=8, was used.

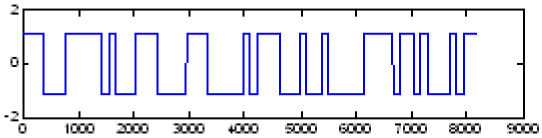

Figure 1. The waveform of the useful component of the input signal.

In the following figure is represented the dependency between the output and input SNRs for different

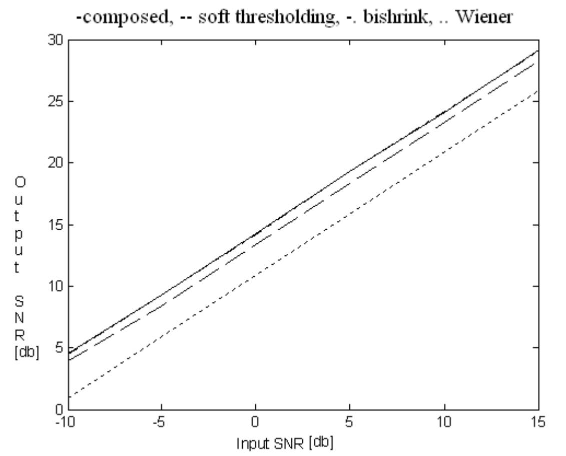

Figure 2. The dependence of the output SNR of the input SNR.

denoising methods. The filter used in the wavelets domain gives the difference. It can be observed that these dependencies are linear. The curves describing the bishrink filter and the composed bishrink filter are superposed. These filters give the better results. These are superiors to the results obtained using the soft-thresholding filter. The poor results are obtained using the Wiener filter. Finally, a comparison between the use of an adapted filter and a denoising system, into a communication application is discussed. Each of the two systems are connected at the output of a communication channel, that adds a zero mean white noise to the data sequence, which beginning is represented in figure 1. The first system is a filter adapted to a rectangle, having a duration equal with *b*. Six experiments are made, with different noise variances. At the output of the investigated system (adapted filter or denoising system), an ideal sampling system is connected. Three hypotheses, concerning the synchronization, are used. The first hypothesis supposes a perfect synchronization. The second hypothesis accepts a little loss of synchronization (the sampling moments are delayed with $3 \cdot b/8$) and the third hypothesis accepts a more important loss of synchronization (the sampling moments are delayed with $b/2$). The output of the sampling system is connected to the input of a comparator. The output of this comparator represents the output of the simulated receiving unit. The denoising system uses a soft thresholding filter when the input SNR is inferior to –16.65 dB and a composed bishrink when the input SNR is superior to –16.65 dB. The better result is obtained with the adapted filter with perfect synchronization (ad.filt.perf.sync, the continuous line in figure 3). For the other hypotheses, an analysis, taking into account the value of the input SNR must be made. The synchronization losses do not affect the performances of the denoising system (denois.perf.syn, the dashed line in figure 3, denois.part.sync.1 and denois.part.sync. 2) but affect the performances of the adapted filter. In the second hypothesis, the denoising system is better than the

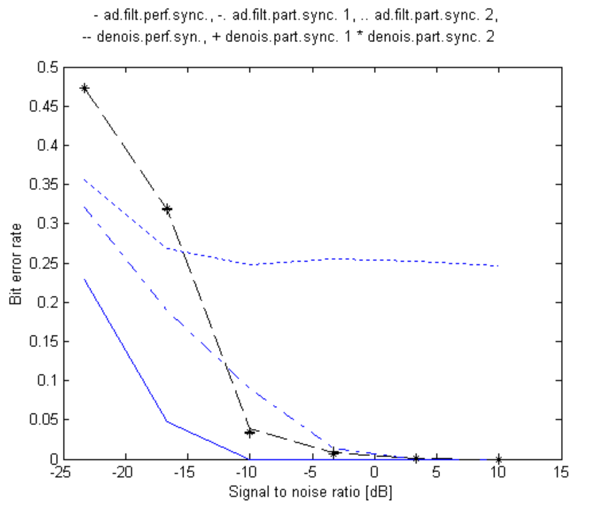

Figure 3. A comparison between the use of an adapted filter and a denoising system, in a base-band communication application.

adapted filter (ad.filt.part.sync 1, the dash dot line in figure 3) for input SNRs superior to –10 dB. Also, in the third hypothesis, the denoising system is superior to the adapted filter (ad.filt.part.sync 2, the dotted line in figure 3), for input SNRs superior to –10.47 dB. Practically in this case the adapted filter cannot be used.

VI. CONCLUSION

In this paper is proposed a new denoising method based on the use of the composed bishrink filter in the wavelets domain. This method takes into account also the statistics of the useful part of the input signal. That makes that this method to perform better than the denoising method using the soft thresholding filter for input SNRs superior to –10 dB. It can be used in communications, replacing the adapted filter, when the synchronization is difficult.

REFERENCES

[1] K. O. Egiazarian, V. P. Melnik, V. V. Lukin, J. T. Astola, “Local transform-based denoising of radar image processing”, Nonlinear Image Processing and Pattern Analysis XII, Edward R. Dougherty, Jaakko T. Astola, Editors, Proceedings of SPIE vol. 4304, 2001.

[2] D. L. Donoho, "De-noising by Soft Thresholding", Technical Report no.409, Stanford University, December 1992.

[3] M. Borda, D. Isar, "Whitening with wavelets", Proceedings of ECCTD'97 Conference, Budapest, August 1997.

[4] A. Isar, A. Cubițchi, Miranda Naforniță, “Algorithmes et techniques de compression”, Editura “Orizonturi universitare”, Timisoara, 2002.

[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] H. Zhang, A. Nostratinia, R. O. Wells Jr., Image Denoising via Wavelet-Domain Spatially Adaptive FIR Wiener Filtering, IEEE ICASP, Istanbul, June 2000, vol. 5, 2179-2182.