WAVELET SHRINKAGE WITH CORRELATED WAVELET COEFFICIENTS

Z. Azimifar *P. Fieguth* *E. Jernigan*

Department of Systems Design Engineering
University of Waterloo
Waterloo, Ontario, Canada, N2L-3G1

ABSTRACT

This paper investigates the statistical characterization of multiscale wavelet coefficients corresponding to random signals and images. Virtually all approaches to wavelet shrinkage model the wavelet coefficients as independent; we challenge that assumption and demonstrate several cases where substantial correlations may be present in the wavelet domain. In particular, the correlation between scales can be surprisingly substantial, even for pixels separated by several scales. Our goal, initiated in this paper, is to develop an efficient random field model describing these statistical correlations, and demonstrate its effectiveness in the context of Bayesian wavelet shrinkage for signal and image denoising.

1. INTRODUCTION

This paper investigates the correlation structure of wavelet coefficients corresponding to various random fields. Specifically, we are interested in studying Bayesian methods of wavelet shrinkage for the purposes of image denoising.

There has been a tremendous amount of activity and interest in the applications of wavelet analyses [1, 2] to signals, in particular methods of wavelet thresholding and shrinkage [3, 4] for the removal of additive noise from corrupted signals and images. The fundamental motivation behind these approaches is that the statistics of many real-world signals, when wavelet transformed, are substantially simplified. However, virtually all Bayesian shrinkage methods model the wavelet coefficients as *independent* Gaussian random variables, an assumption which we challenge.

Modeling wavelet coefficients as independent is sensibly motivated by the fact that the wavelet transform is an effective whitener for a wide variety of random processes [4]. However the wavelet transform is not a perfect whitener — that is, the wavelet coefficients normally *do* possess some degree of correlation, both within and across scales. This fact is not unknown in the literature, as established by Flandrin [5] for fractional Brownian motion, by recent works using hidden Markov models [2] and steerable pyramid correlation models [6] that similarly-sized coefficients tend to cascade along the branches of the wavelet tree. However by and large such wavelet coefficient correlations have been ignored.

This paper concentrates on studying the *exact* within-scale and across-scale statistical dependencies of the wavelet coefficients for a variety of wavelets and random fields, with examples provided for both 1-D and 2-D signals. The results show the whitening effect of the wavelet transform to be quite clear – even for highly correlated spatial processes the coefficients *within* a scale can be nearly unrelated, however the correlation *between* scales is surprisingly substantial, even for pixels separated by several scales. Some of the most interesting results with additional details are highlighted.

2. DEVELOPMENT

The wavelet transform Wf of a signal f is a process in which low and high frequency components of f are represented by separate sets of coefficients, namely the approximation $\{a_J\}$ and the detail $\{d_j\}, 1 \leq j \leq J$. If, as usual, we define the linear operators H and G as high- and low-pass filters respectively, then clearly the coefficient vectors may be recursively computed in scale as

$$a_j = H_{j+1} a_{j+1}, \tag{1}$$

$$d_j = G_{j+1} a_{j+1}. \tag{2}$$

Having defined the sets of $\{a_j\}$ and $\{d_j\}$, we can recursively calculate the within-scale and across-scale covariances from the covariance Σ_{a_j,a_j} at the finest scale $j = 1$ as follows:

$$\begin{aligned} \Sigma_{d_{j+1},d_{j+1}} &= G_j \Sigma_{a_j,a_j} G_j^T \\ \Sigma_{a_{j+1},a_{j+1}} &= H_j \Sigma_{a_j,a_j} H_j^T \\ \Sigma_{a_{j+1},d_{j+1}} &= H_j \Sigma_{a_j,a_j} G_j^T \end{aligned} \tag{3}$$

etc.

The support of the Natural Science & Engineering Research Council of Canada and Communications & Information Technology Ontario are acknowledged.

0-7803-6725-1/01/$10.00 ©2001 IEEE

Vannucci and Corradi [7] proposed a link between the 2-D wavelet transform of Σ_{a_1,a_1} and the overall covariance matrix of 1-D wavelet coefficients, Σ_{Wf}. Having this well-defined variance-covariance structure of the wavelet coefficients, one can assess the extent of *correlation* between the coefficients at the same scale or different resolutions.

An exponential correlation structure is common for real images and remotely-sensed fields, so we will assume that the second-order statistics of the finest-scale signal f are given by $f \sim (0, \Sigma_f)$, that is, f is zero mean and has covariance structure

$$\Sigma_f = cov(lag) = \sigma_f^2 \exp(-\frac{lag}{\tau}) \quad (4)$$

with parameter τ controlling the correlation length between two pixels. Our chosen distribution for now has constant correlation length and is spatially stationary; this assumption is for convenience only and is not fundamental to our analysis.

With the covariance structure Σ_f determined, we transform it into the wavelet domain by computing the wavelet kernel W, containing all translated and dilated versions of the selected wavelet basis functions. The covariance structure of the wavelet-decomposed signal is then

$$\Sigma_{Wf} = W \Sigma_f W^T. \quad (5)$$

3. RESULTS - 1D

After generating the 1-D wavelet transform kernel W we find Σ_{Wf} as in (5). For convenience in understanding the results, the resulting variances are normalized, so that the inter-coefficient relationships are measured as correlation coefficients $-1 \leq \rho \leq 1$. For the purpose of this paper we show the results of simulation with the piecewise linear family of Haar basis functions. Our simulations with other bases functions, such as the Daubechies wavelets which are more regular, exhibit a stronger decorrelation effect within scale, nevertheless the qualitative structure is similar, and the across-scale correlations are no less significant.

The resulting correlation is a block matrix, with the block diagonals showing the within-scale autocorrelations and off-diagonal blocks containing the cross-correlations between $\{a_J\}$ and $\{d_j\}, 1 \leq j \leq J$, in different resolutions.

Figure 1 summarizes the magnitudes of the correlation coefficients between a typical detail coefficient and its spatially local neighbors, both within the same scale and across scales. We see most clearly that the within-scale correlations tends to decay very quickly, consistent with the understanding that the wavelet transform is whitening the original signal which supports the notion of fast, the dependencies across different resolutions surprisingly remain strong, even for coefficients located several scales away from each other. This result confirms that although the wavelet coefficients

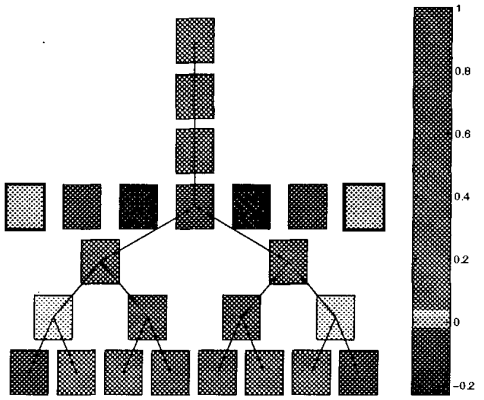

Fig. 1. The extent of correlation between a typical coefficient of 1-D WT at scale $s = j$ and its adjacent coefficients within the same scale and across several resolutions towards both parents and children.

are expected to be uncorrelated, there exist cases for which the correlation can be quite significant.

4. RESULTS - 2D

We extended our joint-statistical study of the 1-D wavelet coefficients to 2-D wavelet transform. Since the size of the covariance matrix for even a small $n \times n$ image increases dramatically to $n^2 \times n^2$, the empirical results were limited to considering the correlation structure of 32×32 images.

Figure 3 illustrates the correlation structure of 2-D wavelet coefficients of a 3-level wavelet decomposition. The main diagonal blocks show the autocorrelation of coefficients located at the same scale and at the horizontal, vertical and diagonal orientations respectively. Due to the column-wise 2-D to 1-D data stacking, large magnitude autocorrelations of the vertical coefficients (labeled as v) tend to concentrate near the main diagonal, whereas those of the horizontal coefficients (labeled as h) are distributed on the diagonals n pixels apart.

Figure 2 presents the two-dimensional parallel of Figure 1, showing the correlation pattern for a typical horizontal detail coefficient. It is significant to notice that regardless of the orientation, the large magnitude correlations at any sub-band are basically arranged in lines following the orientation of the sub-band. The strong horizontal correlation of the horizontally aligned coefficients is apparent, even for pairs of coefficients across different resolutions, as are the inter-scale correlations. By symmetry we can expect similar correlation structures for the vertical coefficients, but clearly in the vertical direction.

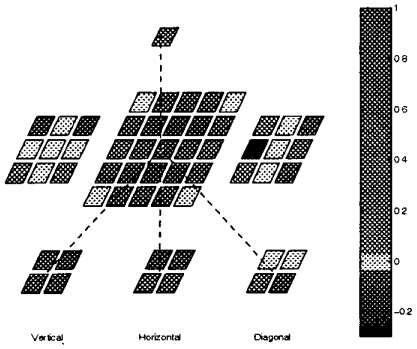

Fig. 2. Summary of correlation in 2-D wavelet domain between a horizontal coefficient and its spatially local neighbors at the same scale, but different orientations and across scales but the same orientation.

5. WAVELET-BASED BAYESIAN LEAST SQUARE THRESHOLDING

Our primary research goal is the development of a multiscale-based Bayesian denoising algorithm, which explicitly takes into account some prior model of the *true* correlation structure exhibited by the wavelet coefficients.

As is clear from Figures 1 and 2, there is a clear locality to the correlation structure and the covariance matrix is highly structured. In order to consider these statistical dependencies we implement a method that estimates the original coefficients by explicit use of wavelet covariance structure. Due to the linearity of the wavelet transform, Bayesian Least Square (BLS) method which explicitly takes into account the wavelet coefficients covariance structure is considered and is given in (6).

$$\hat{\mathbf{f}} = \Sigma_f^T(\Sigma_f + R)^{-1}\mathbf{g} \tag{6}$$

The goal is to estimate $\mathbf{f} \sim N(0, \Sigma_f)$ from noisy observation $\mathbf{g}$, where additive noise $\mathbf{v} \sim N(0, R)$ is uncorrelated with original data $\mathbf{f}$.

We only need to re-consider (6) in the wavelet domain. It is clear that wavelet transform is a linear operation, hence, BLS still applies in that domain. To transform the covariance structure we only need to substitute (5) into (6). We would have the orthogonal wavelet transform of BLS method as

$$\hat{f} = W^{-1}[W\Sigma_f W^T(W\Sigma_f W^T + WRW^T)^{-1}Wg] \tag{7}$$

We considered an ensemble of parameterized Gaussian Markov Random Fields (GMRF) with predefined covariance structure. In order to perform appropriate comparisons and also to emphasize on importance of considering wavelet coefficients correlation – within-scale and across-scale, three different structures of $W\Sigma_f W^T$ were considered in BLS estimate:

GMRF	Thin-plate	Membrane	Tree	Grass
noisy(g)	3.4	4.9	2.29	2.55
WT-Wiener	2.3	4.3	1.96	2.17
WT-BLS1	1.9	3.8	1.54	1.73
WT-BLS2	1.7	3.75	1.22	1.43
WT-BLS3	*1.4*	*3.6*	*1.15*	*1.12*
SBLS	*1.4*	*3.6*	*1.15*	*1.12*

Table 1. MSE measure of noisy observation g and denoised images obtained by BLS method and different covariance structures, compared with spatial domain BLS and wavelet-based Wiener thresholding

1. WT-BLS1: Decorrelated coefficients

   First, only diagonal elements of $W\Sigma_f W^T$ were used, while all other elements were replaced with zero. In this case, indeed, wavelet transform was assumed as a whitening process, i.e., all coefficients were considered as independent.

2. WT-BLS2: Within-scale correlated coefficients

   Next, a partial correlation was taken into estimation process. In fact, the coefficients at every scale and for each orientation were considered to be correlated. However, across-scale correlation was ignored, i.e., the parent-child relation was disregarded.

3. WT-BLS3: Within-scale and across-scale correlated coefficients

   Finally, the complete covariance structure as shown in Figure 3 was used by our wavelet-based BLS algorithm.

The wavelet-based BLS algorithm was implemented with each of these three approaches. As expected, the third method, i.e., full covariance structure outperforms other two alternatives. Table 1 illustrates the MSE of denoised image for each case obtained for four different GMRF's. The equivalence between wavelet-based BLS3 and that of spatial-based BLS3 (SBLS) indicates the optimality of considering the complete wavelet coefficients correlation structure. It is also clear that our BLS3 outperforms the wavelet-based Wiener thresholding [9].

As clear from Figure 4, WT-BLS3 which uses full covariance matrix outperforms other methods in psycho-visual appearance of denoised image. Since our BLS algorithm modifies all frequency components of image, the result looks

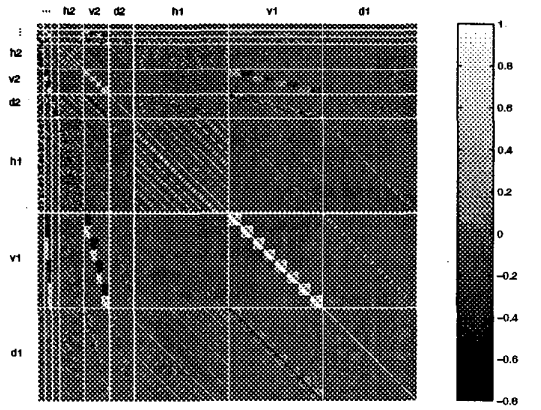

Fig. 3. Correlation structure of an exponentially correlated image decomposed into the 3-level wavelet domain. The main diagonal blocks show autocorrelation of coefficients located at the same scale and at the horizontal, vertical and diagonal orientations respectively, whereas the off diagonal blocks illustrate cross-correlations within or across scales.

blurred. This smoothness can be reduced by leaving the approximate coefficients unchanged, as most of wavelet-based denoising algorithms do.

6. RESEARCH DIRECTIONS

According to our achievements of statistical dependencies between the wavelet coefficients we propose to model the wavelet coefficients not as independent, but as governed by a Markov random field. There are several directions and challenges associated with this kind of undertaking:

- By coupling the wavelet coefficients, the shrinkage problem is complicated considerably, in that the processing of the wavelet coefficients now depends on all others, in precisely the same way that inverting a banded matrix is much harder than a diagonal one.

- There exists past literature on the use of wavelets for the preconditioning of linear systems problems [8]. Such preconditioning is mathematically very similar to the wavelet change of basis in wavelet shrinkage, and may have insights to offer.

- Since correlations are present both within and across scales, a random field model for the wavelet coefficients with itself needs to be hierarchical. The development of Markov random field methods on hierarchies has some past literature, but is still relatively new.

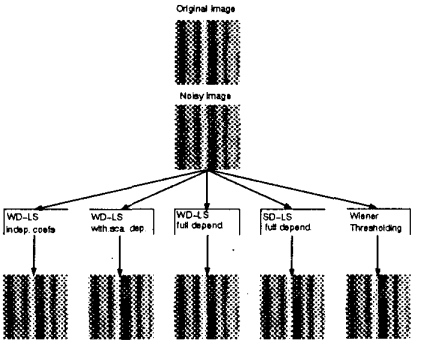

Fig. 4. Denoised image by BLS method with variety of covariance structures compared with wavelet-based Wiener filtering. Fifth-ordered GMRF (tree texture).

7. REFERENCES

[1] Mallat S. G., "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE trans. on PAMI*, vol. 11, pp. 674–693, 1989.

[2] Crouse M. S., Nowak R. D., and Baraniuk R. G., "Wavelet-based statistical signal processing using hidden markov models," *IEEE trans. on SP*, vol. 46, pp. 886–902, 1998.

[3] Donoho D.L., "De-noising by soft-thresholding," *IEEE tras. on IT*, vol. 41, pp. 613–627, 1995.

[4] Chang S. G., Yu B., and Vetterli M., "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE trans. on IP*, vol. 9, pp. 1522–1531, 2000.

[5] Flandrin P., "Wavelet analysis and synthesis of fractional brownian motion," *IEEE trans. on IT*, vol. 38, pp. 910–916, 1992.

[6] Portilla J. and Simoncelli E., "Image denoising via adjustment of wavelet coefficient magnitude correlation," *Proceedings of the 7th ICIP, Canada.*, 2000.

[7] Vannucci M. and Corradi F., "Covariance structure of wavelet coefficients: theory and models in a bayesian perspective," *J. R. Statis. Soc. B*, vol. 61, pp. 971–986, 1999.

[8] Yaou M.H. and Chang W.T., "Fast surface interpolation using multiresolution wavelet transform," *IEEE tras. on PAMI*, vol. 16, pp. 673–688, 1994.

[9] Nowak R. D., "Wavelet-based rician noise removal for magnetic resonance imaging," *IEEE trans. on IP*, vol. 8, pp. 1408–1419, 1999.