MODELING THE CORRELATION STRUCTURE OF IMAGES IN THE WAVELET DOMAIN

Z. Azimifar, P. Fieguth, E. Jernigan

Department of Systems Design Engineering
University of Waterloo
Waterloo, Ontario, Canada, N2L-3G1
azimifar@uwaterloo.ca

ABSTRACT

In this paper we investigate the correlation structure of the wavelet coefficients corresponding to random fields. The context of this work is the study of Bayesian approaches to wavelet shrinkage for the purposes of image denoising. This paper concentrates on both within-scale and across-scale statistical dependencies 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 particular highly correlated spatial processes the within-scale correlation decays exponentially fast, however the correlation between scales is surprisingly substantial, even for separations several scales apart. Our goal, initiated in this paper, is the development of an efficient random field model, describing these statistical correlations, and the demonstration of its effectiveness in the context of Bayesian wavelet shrinkage for signal and image denoising.

1. INTRODUCTION

The past few years have seen considerable activity in multiresolution wavelet analysis [1, 2] of signals and images, in particular methods of wavelet thresholding and shrinkage [3, 4] for the removal of additive noise from corrupted signals. However, virtually all Bayesian shrinkage methods model the wavelet coefficients as *independent* and identically distributed random variables.

The models based on independent wavelet coefficients are 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 between and across scales. This fact is not unknown in the literature, as established by Flandrin [5] for fractional Brownian motion, and recent work in wavelet denoising with a hidden Markov model [2], however, by and large the correlations have been ignored.

In this paper, we propose to challenge the assumption of the wavelet transform as an always perfect whitening operation. To make our prior model of wavelet coefficients more accurate and to improve significantly its performance both in peak SNR and visual metrics it is necessary to design a statistical model for the correlation structure of wavelet coefficients of real images.

In order to propose a strategy for a modified wavelet denoising algorithm which incorporates some understanding of wavelet correlations we need to determine two things:

1. The wavelet coefficient correlations;
2. A method for taking advantage of the known correlations.

This paper concentrates on the former issue, a straightforward Monte-Carlo study of the *within-scale* and *across-scale* statistical dependencies of the wavelet coefficients for a variety of wavelets and random fields. The results show that the within-scale correlation decays very fast, whereas the inter-scale correlation remains substantial, even for separations of several scales apart.

2. WAVELET THRESHOLDING

An image is often degraded by noise in different stages, such as in discretization or transmission. The goal of cleaning a noisy image is to remove the corrupting noise while retaining the important features of the image. Traditionally, image denoising is achieved by classical linear Wiener filtering [6]. However, a tremendous interest has recently emerged on using nonlinear *wavelet thresholding* or *shrinkage* techniques [3, 4, 7] for removing additive white Gaussian noise from the wavelet transform of a corrupted image.

Let $\{y(m,n)\}$ represent the measurement of the original signal $x(m,n)$ corrupted by additive noise $\varepsilon(m,n)$

$$y(m,n) = x(m,n) + \varepsilon(m,n), \quad m,n = 1,\cdots,N \quad (1)$$

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

Fig. 1. Image denoising through wavelet thresholding.

where the $\{\varepsilon(m,n)\}$ are *iid* normal $N(0,\sigma^2)$ and independent of $\{x(m,n)\}$. The goal of denoising is to remove the noise from the observations $\{y(m,n)\}$ by computing estimates $\{\hat{x}(m,n)\}$ which satisfy aesthetic psychovisual and/or mean square error (MSE) criteria,

$$MSE(\hat{x}) = \frac{1}{N^2}\sum_{m,n=1}^{N}(\hat{x}(m,n) - x(m,n))^2 \quad (2)$$

Let $Y = W\mathbf{y}$ denote the matrix of wavelet coefficients of $\{y(m,n)\}$, where W stands for the 2-D dyadic orthogonal wavelet transform process and $\mathbf{y}$ denotes the observed image y in a long vector format. Similarly, the 2-D wavelet transform of $X = W\mathbf{x}$ and $V = W\varepsilon$ satisfy (1) as

$$Y_j = X_j + V_j, \quad 1 \le j \le J \quad (3)$$

where J indicates the coarsest resolution level. The orthogonality of the wavelet transform implies that $\{V_j(i,k)\}, 1 \le i,k \le N/2^j$ are also *iid* normal $N(0,\sigma^2)$.

The wavelet thresholding operation $\mathbf{T}$ filters each detail wavelet coefficient $\{Y_j(i,k)\}$ with a predefined threshold λ to obtain $\{\hat{X}_j(i,k)\}$.

$$\hat{X}_j = \mathbf{T}(Y_j,\lambda), \quad \hat{\mathbf{x}} = W^{-1}[\hat{X}_j] \quad (4)$$

Eq. (4) indicates that the inverse wavelet transform W^{-1} is then applied to the denoised estimate to achieve $\hat{\mathbf{x}}$. The complete process of wavelet thresholding is shown in Fig. 1.

While the idea of thresholding is simple and effective, finding a good threshold is not an easy task. Due to the mathematical property of the wavelet transform, i.e. compact support, only a few large wavelet coefficients really contain information about the original image while a large number of small coefficients are attributed to the noise [6]. This reasonable assumption has motivated various wavelet thresholding algorithms which are classified into three major categories:

- Universal thresholding [6]
- Data-adaptive thresholding [4]
- Bayesian thresholding [2, 3]

In the statistical Bayesian literature, many works have been concentrated on deriving the best threshold based on the prior models such as Laplacian [4] and mixture of Gaussian [3]. These prior models assume the wavelet coefficients are *independent* and identically distributed. They are thus completely determined by the marginal statistics of the residual coefficients. On the other hand, Crouse et. al [2] model the coefficients as a Hidden Markov model by considering only two hidden states for each coefficients. Portilla and Simoncelli [7] also model the coefficients by capturing their dependencies within-scale and with the parent scale, only.

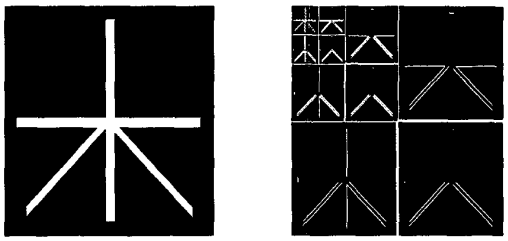

Fig. 2. Three-level wavelet decomposition of a synthesized image. White pixels indicate large magnitude coefficients, and black pixels carry small magnitude.

The major reason for independent models is based on the localization property, i.e., the parsimonious representation, plus the interpretation of the wavelet transform as a decorrelator that attempts to make each wavelet coefficient statistically independent of all others. However, the wavelet transform cannot completely decorrelate real valued signals and a residual dependency structure usually remains between the coefficients. Although the family of wavelet bases constitutes an orthogonal system, there is *a priori* no reason for the wavelet coefficients to be uncorrelated [5].

In this work, we assume the wavelet coefficients, in general, to be correlated in both space and scale. This fact was observed for non-stationary Brownian motion processes [5] and for real images [2, 7]. It would be, however, interesting to evaluate how strong this correlation is, or even to point out special cases for which decorrelation could be achieved.

3. JOINT STATISTICAL MODEL OF WAVELET COEFFICIENTS

3.1. Joint Histogram of Wavelet Coefficients

Our experiments with wavelet coefficients of a group of real images revealed that along the cascade of several scales there exists a persistency [2] between the magnitude of parent and child coefficients. This important property implies the statistical dependencies along the different resolutions. This fact becomes more obvious from Fig. 2 which shows the magnitudes of coefficients in a three-level separable wavelet decomposition of a synthesized image. Note that large-magnitude coefficients tend to occur near each other within subbands, and also occur at the same relative spatial loca-

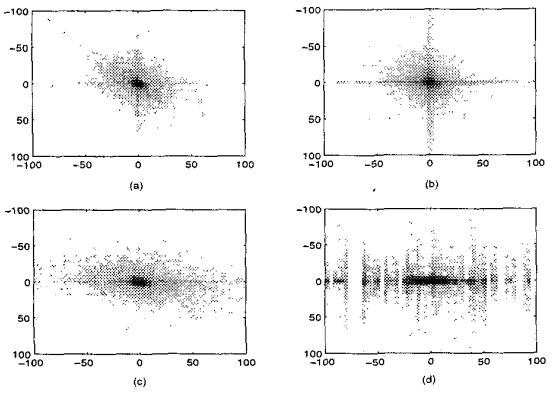

Fig. 3. Joint histograms of wavelet coefficients:
(a) Histogram of coefficients $d_{j,k}$ and $d_{j,k+1}$;
(b) Histogram of coefficients $d_{j,k}$ and $d_{j,k+5}$;
(c) Histogram of coefficients $d_{j,k}$ and $d_{j+1,k}$;
(d) Histogram of coefficients $d_{j,k}$ and $d_{j+3,k}$.

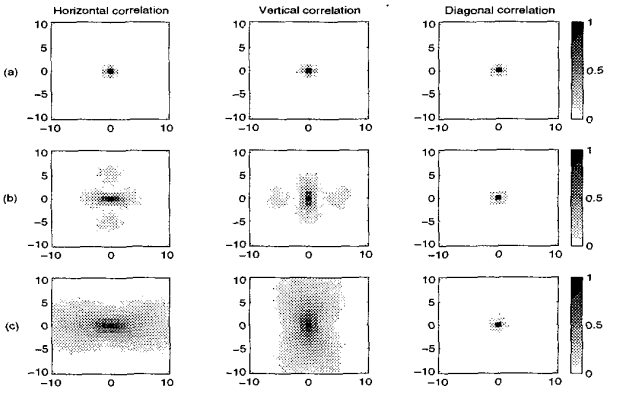

Fig. 4. The within scale length of correlation for 2-D Haar wavelet coefficients when the spatial correlation length is: (a) 1 pixel, (b) 10 pixels, (c) 30 pixels.

tions in subbands at adjacent scales, and orientations.

To make this fact more explicit, Fig. 3 shows joint histograms of four different pairs of coefficients. Here, ergodicity and stationarity are assumed, i.e.,

$$E[x_{i,j}x_{i+\Delta i,j+\Delta j}] = E[x_{0,0}x_{\Delta i,\Delta j}], \quad 1 \le i,j \le N \quad (5)$$

in order to consider the joint histogram of pairs of coefficients. Each plot is obtained by counting the number of corresponding pairs of coefficients whose magnitude belongs to a predefined interval. The plots (a) and (b) in Fig. 3 illustrate that at a particular resolution j the adjacent coefficients $d_{j,k}$ and $d_{j,k+1}$ produce joint histograms that exhibit the correlation between coefficients, whereas the far away coefficients, e.g. $d_{j,k}$ and $d_{j,k+5}$, produce a nearly circular histogram. Fig. 3(c,d) shows the coefficients joint histograms in the nearby scales $d_{j,k}$ and $d_{j+1,k}$ are clearly extended along the axes, which indicates strong dependency along the resolutions. Indeed, these plots show the underlying joint density of the wavelet coefficients.

The preliminary observations of the joint statistics of the wavelet coefficients have revealed that as the further distance coefficients (either in spatial positions or across different scales) are chosen, the dependencies become weaker. The conditional dependency given a square of neighborhood within or across scales suggests investigating the Markov random field models of wavelet coefficients.

3.2. Monte-Carlo Study of Wavelet Correlation Coefficients

First, we present the results of the straightforward Monte-Carlo analysis, where the finest-scale image is a sample of a parameterized random field. With the associated ensemble of wavelet-transformed images we were able, albeit only approximately, to highlight some of the significant residual correlations between coefficients within and across scales.

A. *Random field generation*

The statistical study of the wavelet coefficients was started by generating an ensemble of parameterized random fields with a spatially stationary assumption. Because of the stationary property, the correlation structure of any $N \times N$ random field takes the same form as Eq. 5. Therefore the autocorrelation structure of the zero-mean random field would be

$$\Sigma_{i,j} = E[x_{0,0}x_{i,j}], \quad 1 \le i,j \le N \quad (6)$$

The covariance matrix is toroidally stationary, i.e., a circulant matrix, which can be diagonalized by the 2-D FFT. A Gaussian random field $\mathbf{X}$ is then synthesized as

$$\mathbf{X} = FFT^{-1}\{\sqrt{FFT(\Sigma)} \cdot FFT(Q)\} \quad (7)$$

Where $sqrt(\cdot)$ and $\cdot$ are element-by-element operations and Q is a matrix of unit variance Gaussian random variables with the same size as Σ

$$Q \sim N(0, I) \quad (8)$$

B. *Correlation coefficients*

A variety of sample path ensembles of both small and large correlation lengths were transformed into the wavelet domain. For each case the within-scale sample correlation coefficients were calculated for a local spatial neighborhood at the same orientation, i.e., horizontal, vertical or diagonal direction. For convenience in understanding the results,

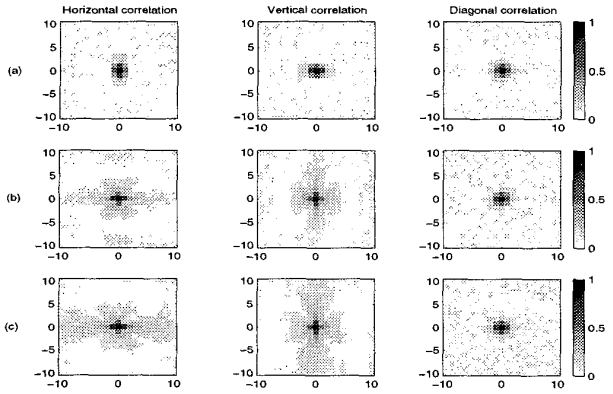

Fig. 5. The within scale length of correlation for 4-tap Daubechies wavelet coefficients when spatial correlation length is: (a) 1 pixel, (b) 10 pixels, (c) 30 pixels.

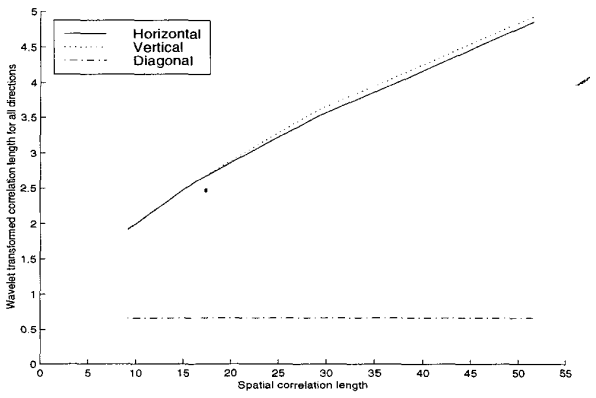

Fig. 6. A plot of within-scale correlation length for Haar wavelet coefficients in all directions vs. spatial correlation length.

the resulting variances are normalized, so that the inter-coefficient relationships are measured as correlation coefficient

$$\rho = \frac{E[(\mathbf{x} - \mu_x)(\mathbf{y} - \mu_y)]}{\sigma_x \sigma_y} \tag{9}$$

where $-1 \leq \rho \leq 1$. Indeed, $|\rho| = 1$ shows the coefficients are completely correlated and the value $\rho = 0$ indicates totally decorrelation between two wavelet coefficients. Fig. 4 summarizes the extent of the correlation for a typical coefficient of the Haar wavelet basis in the horizontal, vertical or diagonal channels. It also shows the change of the correlation length, from Fig. 4(a) to Fig. 4(c), at a typical resolution when the spatial correlation length increases. This simulation was repeated for the commonly used Daubechies wavelets. Fig. 5 illustrates the within-scale correlation length for the 4-tap Daubechies wavelet, albeit less significant than the results achieved for the Haar wavelet.

Fig. 6 plots the within scale correlation length in the Haar wavelet domain for all three orientations vs. the length of the spatial correlation. Although the whitening effect of the wavelet transform is quite clear for the within scale diagonal coefficients, the coefficients in the horizontal and vertical directions exhibit a residual relation along their orientation which grows, albeit very slowly, with increasing correlation length in the spatial domain. It worths mentioning here, that the decorrelation illustrated for diagonal coefficients is a direct consequence of generated random fields considered in our simulations. If diagonal correlation in the spatial domain increases, it is expected that the diagonally aligned wavelet coefficients also exhibit strong dependency. Fig. 7 also highlights the correlations between pairs of horizontally aligned coefficients at four different resolutions. The increasing trend of correlation between the coefficients up to five pixels apart is quite obvious in this plot.

4. DISCUSSIONS

Our primary research goal is the development of a multiscale-based Bayesian denoising algorithm, which explicitly takes into account some prior model, as illustrated in Figures 3 to 5, of the *true* correlation structure exhibited by the wavelet coefficients.

In this paper we have described the joint statistics for ensemble of generated random fields in the wavelet domain. The empirical wavelet coefficient correlations illustrate a substantially more powerful joint model than the traditional Generalized Gaussian models proposed based on the independent assumption for the coefficients. As is demonstrated by these figures, there is a clear locality to the correlation structure, and so we propose to model the wavelet coefficients not as independent, but as governed by a Markov random field. We have started to investigate the use of multiscale MRF models to capture the conditional density based on a neighborhood of wavelet coefficients. Since correlations are present both within and across scales, a random field model for the wavelet coefficients with itself need to be hierarchical. The development of Markov random field methods on hierarchies has some past literature, but is still relatively new.

5. 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 Signal Processing*, vol. 46, pp. 886–902, 1998.

[3] Abramovich F., Sapatinas T., and Silverman B. W., "Wavelet thresholding via a bayesian approach," *J. R. Statis. Soc. B*, vol. 60, pp. 725–749, 1998.

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

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

[6] Mallat S. G., *A wavelet tour of signal processing*, Academic Press, 1999.

[7] Portilla J. and Simoncelli E., "Image denoising via adjustment of wavelet coefficient magnitude correlation," *Proceedings of the 7th International Conference on Image Processing, Vancouver, BC, Canada.*, 2000.

[8] 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.

[9] Romberg J., Choi H., Baraniuk R., and Kingsbury N., "Multiscale classification using complex wavelets and hidden markov tree models," *Proceedings of the 7th International Conference on Image Processing, Vancouver, BC, Canada.*, 2000.

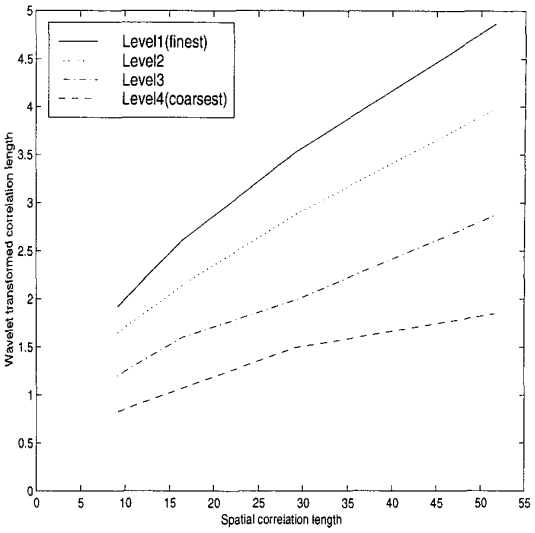

Fig. 7. Within-scale correlation length for horizontal Haar wavelet coefficients at different scales vs. spatial correlation length.