Hierarchical Multiscale Modeling
of Wavelet-Based Correlations

Zohreh Azimifar, Paul Fieguth, and Ed Jernigan⋆

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

Abstract. This paper presents a multiscale-based analysis of the statistical dependencies between the wavelet coefficients of random fields. In particular, in contrast to common decorrelated-coefficient models, we find that the correlation between wavelet scales can be surprisingly substantial, even across several scales. In this paper we investigate eight possible choices of statistical-interaction models, from trivial models to wavelet-based hierarchical Markov stochastic processes. Finally, the importance of our statistical approach is examined in the context of Bayesian estimation.

1 Introduction

This paper presents a hierarchical multiscale (MS) model to describe the statistical dependencies between the wavelet coefficients as a first-order Markov process. The model is premised on the fact that, regardless of their spatial locations, wavelet coefficients are highly correlated across scales, even those separated by several scales. The virtue of this model is its ability to capture coefficients correlations by concentrating on a very sparse statistical structure. Furthermore, the within-subband coefficients in MS model framework exhibit a clear a Markovian nature.

Our motivation is model-based statistical image processing. That is, we are interested in the statistical manipulation of images, which requires some probabilistic description of the underlying image characteristics. The image pixel interactions in the spatial domain lead to extremely complicated (in particular, highly correlated) statistical structures, which are computationally inconvenient to be used in estimation algorithms. In order to simplify the raw statistics of pixel values, a spatial transformation is considered. The transform is chosen to simplify or nearly decorrelate, as much as possible, the starting statistics, analogous to the preconditioning of complicated linear system problems. The popularity of the wavelet transform (WT) stems from its effectiveness in this task: many operations, such as interpolation, estimation, compression, and denoising are simplified in the wavelet domain, because of its energy compaction and decorrelative properties [1,2].

A conspicuously common assumption is that the WT is a perfect whitener, such that all of the wavelet coefficients are independent, and ideally Gaussian. There is, however, a growing recognition that neither of these assumptions are accurate, nor even adequate for many image processing needs. Indeed, significant dependencies still exist between wavelet coefficients. There have been several recent efforts to study the wavelet statistics; mostly marginal models. Each statistical wavelet model focuses on a certain type of dependencies, in which a relatively simple and tractable model is considered. We classify them into the following two categories:

1. Marginal Models:
    (a) Non-Gaussian, i.e., heavy tail distribution [3],
    (b) Mixture of Gaussians [3],
    (c) Generalized Gaussian distribution [2],
    (d) Bessel functions [4].
2. Joint Models:
    Hidden Markov tree models [1].

In virtually all marginal models currently being used in wavelet shrinkage [2], the coefficients are treated individually and are modelled as independent, i.e., only the diagonal elements of wavelet based covariance matrix are considered. This approach, however, is not optimal in a sense that the WT is not a perfect whitening process.

The latter approach, however, examines the joint statistics of coefficients [5]. Normally an assumption is present that the correlation between coefficients does not exceed the parent-child dependencies, e.g. given the state of its parent, a child is decoupled from the entire wavelet tree [1,6].

It is difficult to study both aspects simultaneously: that is, the development of non-Gaussian joint models with non-trivial neighborhood. The study of independent non-Gaussian models has been thorough; the complementary study, the development of Gaussian joint models, is the focus of this paper. The goal, of course, is the ultimate merging of the two fields. However for the purpose of this paper, we are willing to limit ourselves to simplifying marginal assumptions (Gaussianity) which we know to be incorrect, but which allow us to undertake a correspondingly more sophisticated study of joint models.

The main theme of this paper is, then, to concentrate on studying within and across scale statistical dependencies of the wavelet coefficients for a variety of wavelet basis functions and random fields. These correlations are modelled: from complete independent assumption to full dependency between the wavelet coefficients over the entire resolutions. Since correlations are present both within and across scales, we are interested to model them in a wavelet-based MS framework. Finally, the effectiveness of our statistical-based approach is tested through numerical experiments by exploiting Bayesian estimation technique and we show that adding significant dependencies to the wavelet prior model causes dramatic RMSE reductions.
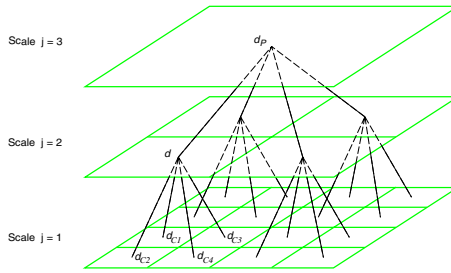
Fig. 1. Illustration of a typical coefficient d along with its parent and children within one wavelet tree subband

2   Discrete Wavelet Transform

The WT of an image f is a process in which the low and high frequency components of f are represented by separate sets of coefficients, namely the approximation $\{a_J\}$ and the detail $\{d_j\}$, $1 \leq j \leq J$, with J denoting the coarsest resolution. If, as usual, we define the linear operators H_j and L_j as high- and low-pass filters respectively, then clearly the coefficient vectors may be recursively computed in scale j

$$a_j = L_{j-1}L_{j-1}a_{j-1}, \qquad d_j^h = H_{j-1}L_{j-1}a_{j-1},$$
$$d_j^v = L_{j-1}H_{j-1}a_{j-1}, \qquad d_j^d = H_{j-1}H_{j-1}a_{j-1} \qquad (1)$$

with $\{d_j^h, d_j^v, d_j^d\}$ denote the horizontal, vertical, and diagonal subbands of the wavelet decomposition at scale j, respectively. The maximum decomposition level for a discrete image with size $n = N \times N$, would be $J = log_2 N$, with $n/4^j$ detail coefficients in every subband at scale j.

Figure 1 illustrates a natural tree of wavelet subbands. Each wavelet coefficient d is shown as a node with d_p as its parent and $\{d_{ci}\}$ $1 \leq i \leq 4$, as the set of its four children, which represent information about this node at the next finer scale. As the scale j decreases, the children add finer and finer details into the spatial regions occupied by their ancestors [1].

2.1   Basic Notations of Wavelet Image Modeling

In order to perform a precise assessment of correlation between the wavelet coefficients of the finest-scale image $f \sim (0, \Sigma_f)$, we consider a variety of prior models based on Gaussian Markov random field (GMRF) covariance structures. The chosen priors, shown in Figure 2 are the tree-bark and thin-plate models. They are spatially stationary, an assumption for convenience only and is not fundamental to our analysis.

The selected covariance structure Σ_f is transformed into the wavelet domain by computing the 2-D wavelet transform W, containing all translated and dilated
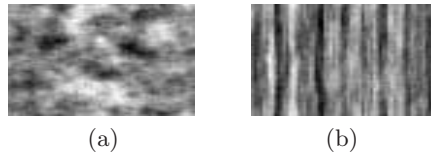
(a)   (b)

Fig. 2. Two GMRF models used in our simulations: (a) Thin-plate, (b) Treebark texture

versions of the selected wavelet basis functions:

$$\Sigma_{Wf} = W \Sigma_f W^T \qquad (2)$$

where we limit our attention to the set of Daubechies basis functions. As more regularity is added to the wavelet function, the within scale decoupling effects increases. Nevertheless, the qualitative structures are similar, and the across-scale correlations are no less significant.

Although in actual data processing we use the covariance matrix, for convenience in understanding the results, the covariance values are normalized, so that the inter-coefficient relationships are measured as correlation coefficients

$$\rho = \frac{E[(d_i - \mu_{d_i})(d_j - \mu_{d_j})]}{\sigma_{d_i}\,\sigma_{d_j}}, \qquad -1 \leq \rho \leq 1 \qquad (3)$$

where d_i and d_j are two typical wavelet coefficients with mean and standard deviation μ_{d_i}, σ_{d_i} and μ_{d_j}, σ_{d_j}, respectively.

In [7] we defined a recursive method to calculate within and across scale covariances for 1-D signals from the covariance Σ_{a_j,a_j} at the finest scale $j = 1$:

$$\Sigma_{d_{j+1},d_{j+1}} = H_j \Sigma_{a_j,a_j} H_j^T, \qquad \Sigma_{a_{j+1},d_{j+1}} = L_j \Sigma_{a_j,a_j} H_j^T, \qquad etc \qquad (4)$$

Having this tool one can easily assess the extent of *correlation* between the coefficients at the same scale or across different resolutions.

Figure 3 illustrates the correlation structure of the 2-D wavelet coefficients of a 4-level wavelet decomposition. Due to dramatic increase in covariance matrix size, the empirical results are limited to considering the correlation structure of 32×32 images. The main diagonal blocks show the autocorrelation of coefficients located at the same scale and orientation. Due to the column-wise 2-D to 1-D data stacking, large magnitude auto-correlations of the vertical coefficients (labeled as v) tend to concentrate near the main diagonal, whereas those of the horizontal coefficients (h) are distributed on the diagonals 32 pixels apart. The off-diagonal blocks contain those cross-correlations of across orientations or across scales.

It is clear that the within-scale correlations tend to decay very quickly, consistent with the understanding that the WT is decoupling the original signal, while the dependencies across different resolutions remain surprisingly strong,
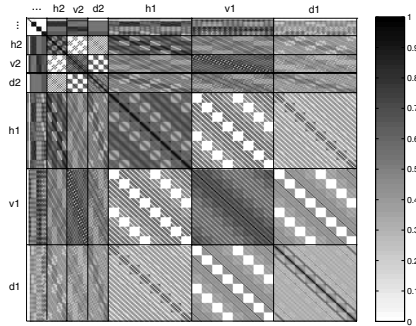
Fig. 3. Scaled values of four-level correlation structure of a thin-plate model decomposed by the Daubechies "db1" wavelet. The main diagonal blocks show autocorrelation of coefficients located at the same scale and orientation, whereas the off-diagonal blocks illustrate cross-correlations across orientations or scales

even for coefficients located several scales apart. This result confirms that although the wavelet coefficients are expected to be decorrelated, there exist cases in which the correlation can be quite significant.

2.2 Numerical Experiments

It is generally infeasible to directly utilize the huge covariance matrix Σ_{Wf} in an estimation process, due to space and time complexities. Our goal is to study the properties of Σ_{Wf} in order to deduce a simple, but still accurate, representation of the underlying correlation model; that is, to construct a new sparse covariance matrix, which contains the most significant information from the prior model. Of course, the study of large covariance matrix is for diagnostic and research purpose; ultimately any practical estimation algorithm will be based on some implicit sparse model of the statistics.

In our experiments the wavelet coefficients are treated in various ways, from complete independence to full dependency among all coefficients over the entire wavelet tree. As is shown in Table 1, eight different cases of adding more features to the new covariance matrix are considered. For each case, except the diagonal case, at least one of the three important neighborhood correlation factors, (*intra-orientation, intra-scale, and inter-scale*) is considered.

Figure 4 visualizes all eight structures obtained from the original correlation matrix. Note that the standard wavelet-based algorithms, in which the coefficients are treated as statistically independent, only consider diagonal entries of the covariance matrix, shown in Figure 4(a). These structures indicate that adding intra-scale correlations increases the structure's density (Figure 4(f)) much more than the inter-scale dependencies (Figure 4(g)). As is evident a large portion of intra-scale correlation values are very close to zero, which says almost nothing about the correlation structure. This fact suggests devising a hierarchical correlation model which keeps its across-scale strength up to several scales,

Table 1. Eight different ways to obtain a new wavelet-based covariance structure which contains a combination of three important neighborhood correlation factors, namely intra-orientation, intra-scale, and inter-scale

Notation	intra orientation	inter orientation	inter scale
diagonal	0	0	0
interorient	0	1	0
interscale	0	0	1
interorient-interscale	0	1	1
inorient	1	0	0
inorient-interorient	1	1	0
inorient-interscale	1	0	1
full	1	1	1

while reducing the within-scale neighborhood to the very close spatial neighbors, i.e., 3×3 spatially located coefficients.

3   Gaussian Multiscale Modeling

As discussed in Section 2.2, the numerical simulations with covariance structure have revealed the importance of taking into consideration a small within-scale correlation range along with a large extent of across-scale dependencies. In order to meet this requirement, there are two alternatives to consider:

1. Imposing models which describe the long range statistical dependencies, such as the full covariance matrix. Such models, however, lead to estimation algorithms that are highly complex and difficult to implement.
2. Proposing a statistical model which tends to approximate the structural correlations over the entire wavelet tree. The advantage of this approach is the existence of estimation techniques which are fast and very easy to implement [8].

Therefore, first-order MS modeling is used to devise an approximation model of the wavelet coefficient correlations. The MS method [8] models each node on the tree as a stochastic process $X(d)$ and recursively describes the interrelation of parent and child as:

$$X(d) = A_d X(d_p) + B_d \nu_d \tag{5}$$

As seen in Figure 1, d represents each node on the tree with its parent denoted as d_p. Here $\nu_d \sim N(0, I)$ is a white noise process and A_d, B_d are parameters to be determined. At the coarsest resolution (root node), the stochastic process $X(d_J)$ is assumed to obey:

$$E[X(d_J)] = 0,$$
$$E[X(d_J)X^T(d_J)] = \mathbf{P}_J \tag{6}$$

(a) diagonal   (b) interorient   (c) interscale   (d) interorient-
                                                            interscale

(e) inorient   (f) inorient-   (g) inorient-   (h) MS model
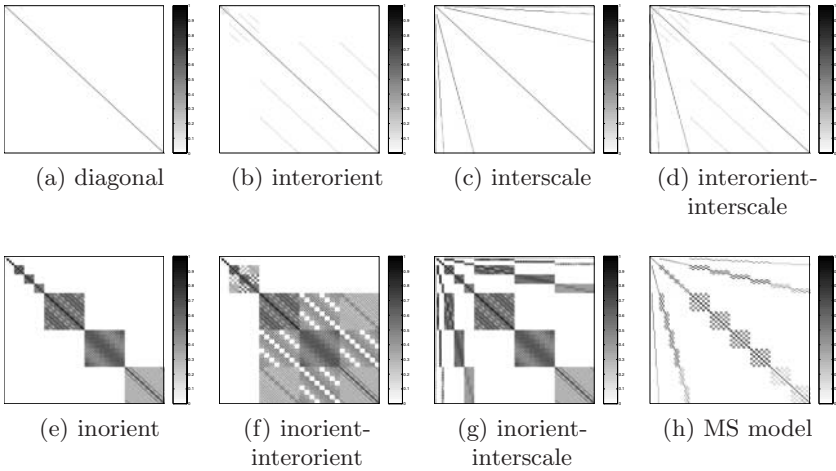   interorient   interscale

Fig. 4. Various correlation structures achieved from the original covariance matrix Σ_{Wf}. (a-g) Seven structures as in Table 1. As within-scale dependencies are considered (f), the structural density increases dramatically. The across scale correlations (g) add significant information, but have less impact on density increment. (h) The correlation structure presented by a multiscale model

Having the initial conditions defined in (6), one can easily calculate the parameters A and B given in the first-order MS model (5). The cross-correlation of each node d and its parent is computed as [8]:

$$\mathbf{P}_{d,d_p} = E[X(d)X^T(d_p)]$$

$$\mathbf{P}_{d,d_p} = A_d \mathbf{P}_{d_p}^T \quad \Longrightarrow \quad A_d = (\mathbf{P}_{d_p}^{-1} \mathbf{P}_{d,d_p})^T$$

$$B_d B_d^T = \mathbf{P}_d - A_d \mathbf{P}_{d_p} A_d^T \tag{7}$$

Figure 4(h) shows the correlation structure obtained by imposing a first-order stochastic process on the original model (Figure 3). Note that inter-scale correlations, even up to distantly separated scales, are well absorbed by this stochastic model. Also observe that the clear locality of neighborhood dependencies demands within-scale Markovanity. These advantages plus the sparse representation of the MS model makes it an elegant tool to capture wavelet-based hierarchical correlations. The corresponding estimation algorithms can thus be implemented with very low computational effort [8].

The accuracy of the MS model can be increased from first-order, (the state of parent is sufficient for a child to be decoupled from all other nodes) to second-order, (the state of its grand-parent is also needed for a node to be independent from the rest of the tree), etc. Another important issue is the number of coefficients that form a node on the wavelet tree. A particular node may contain only a single wavelet coefficient, or two or more coefficients. To illustrate the tradeoff

Table 2. Summary of computational effort required for the MS model to be imposed on a wavelet binary-tree for a 1-D signal of size N. Each number shows the complexity for a combination of MS order and number of coefficients per node

			MS order			
	1	2	3	4	$\cdots$	$(log_2 N)^{th}$-order
	1	-	-	-	$\cdots$	-
Pixels	2	3	-	-	$\cdots$	-
per node	4	6	7	-	-	-
$\downarrow$	8	12	14	21	$\vdots$	$\vdots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
	$N/2$	$\Sigma_{i=1}^2 N/2^i$	$\Sigma_{i=1}^3 N/2^i$	$\cdots$	$\cdots$	$N-1$
Complexity	$O(N)$	$\cdots$	$\sum_{i=0}^{logN-p}[\sum_{j=i}^{i+p-1} 2^j]^3$		$\cdots$	$O(N^3)$

between the order (accuracy) of MS model and computational complexity of the estimation process, a wavelet binary-tree for an exponentially distributed 1-D signal of size N is considered. Various ways of MS modeling, from first-order to $log_2 N$th-order and from single values to vectors of coefficients per node, are examined. Table 2 summarizes time complexity for each MS model. From top left to bottom right the correlation structure becomes more dense while the complexity of even simple estimation algorithms gets harder.

4   Bayesian Wavelet Estimation Approach

A simple estimation algorithm is adopted in this part to evaluate and compare the achieved various statistical structures . To exploit these statistical dependencies we implement a method that estimates the original coefficients by explicit use of wavelet covariance structure. Due to the linearity and orthogonality of the WT, Bayesian Least Square (BLS) method which directly takes into account the covariance structure is:

$$\hat{\mathbf{f}} = \Sigma_f^T(\Sigma_f + R)^{-1}\mathbf{g} \qquad (8)$$

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 decorrelated with original data $\mathbf{f}$. Since the BLS is applied in the wavelet domain, it is necessary to substitute (2) into (8). Then the orthogonal wavelet transform of the BLS method is obtained as:

$$\hat{f} = W^{-1}[W\Sigma_f W^T(W\Sigma_f W^T + WRW^T)^{-1}Wg] \qquad (9)$$

In order to perform appropriate comparisons and also to emphasize the importance of considering wavelet coefficient correlations – within and across scales,
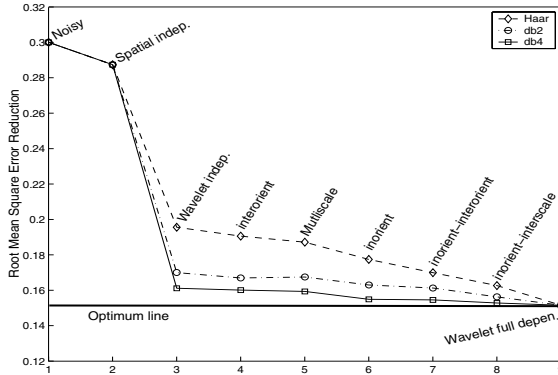
Fig. 5. RMSE measure of noisy observation g and denoised images obtained by BLS method and different covariance structures shown in Figure 4

all structures of Σ_{WF} illustrated in Figure 4 are considered in BLS framework, except those shown in Figure 4(c),(d), due to not being positive definite.

Figure 5 displays the RMSE noise reduction achieved as more correlations are taken into the estimation process. The RMSE performance shows that the more partial correlations (Table 1) are considered, the lower the RMSE. It is extremely important to notice that the rate of RMSE reduction is faster especially if more inter-scale correlations are considered. Larger extent of intra-scale dependencies, however, does not lead to significant RMSE reduction. This fact confirms our earlier discussion of reducing the within-scale neighborhood dependency in our model. As seen in this Figure, the MS-based correlation structure is promising and outperforms the decoupling assumption of the WT, in addition to being a sparse structure of the huge covariance matrix. The MS-based structure with relatively few coefficients vastly reduces the RMSE. Regardless of its well capturing of the across scale dependencies, this model still demands improvements in describing the within scale relations.

5   Conclusions

A multiscale-based analysis of statistical dependencies between the wavelet coefficients was presented. Since correlations are present both within and across scales, wavelet-based hierarchical Markov stochastic processes were proposed and investigated.

The proposed MS model exhibits a sparse locality to the coefficient activities, which results in a dramatic RMSE reduction. The virtue of the model is its ability to capture the most significant statistical information between tree parents and children, however the interrelationship of pixels within a scale is only implicit, and very limited. To complete our development of MS model, we will consider higher local spatial neighboring activities towards a MRF modeling of the wavelet coefficients statistics. The development of MRF methods on

hierarchies has some past literature, but is still relatively new and we are willing to extend this work to the proper MRF modeling of statistical dependencies on spatial neighbors.

References

1. Romberg K., Choi H., and Baraniuk R., "Bayesian tree-structured image modeling using wavelet-domain hidden markov models," *IEEE trans. an IP,* vol. 10, pp. 1056-68,2001.   850, 851, 852
2. Portilla J. and Simoncelli E., "Image denoising via adjustment of wavelet coefficient magnitude correlation," *Proceedings of the 7th ICIP, Cunudu.,* 2000.   850, 851
3. Mumford D. and Huang J., "Statistics of natural images and models," *Proccedings of International Conference an Computer Vision und Pattern Recognition,* 1999. 851
4. Srivastava A., Liu X., and Grenander U., "Analytical models for reduced spectral representations of images," *Proceedings of the 8th ICIP,* 2001.   851
5. E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," *Proceedings of the SPIE 44th Annuul Meeting,* 1999.   851
6. Crouse M. S., Nowak R. D., and Baraniuk R. G., "Wavelet-based statistical signal processing using hidden markov models," *IEEE trans. an Signal Processing,* vol. 46, pp. 886-902,1998.   851
7. Azimifar Z., Fieguth P., and Jemigan E., "Wavelet shrinkage with correlated wavelet coefficients," *Proceedings of the 8th ICIP,* 2001.   853
8. Chou K., Willsky A., and Benveniste A., "Multiscale recuresive estimation, data fusion, and regularization," *IEEE trans. an Automutic Control,* vol. 39, pp. 468-478, 1994.   855, 856