HIERARCHICAL MARKOV MODELS FOR WAVELET-DOMAIN STATISTICS

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

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

ABSTRACT

There is a growing realization that modeling wavelet coefficients as statistically independent may be a poor assumption. Thus, this paper investigates two efficient models for wavelet coefficient coupling. Spatial statistics which are Markov (commonly used for textures and other random imagery) do not preserve their Markov properties in the wavelet domain; that is, the wavelet-domain covariance P_w does not have a sparse inverse. The main theme of this work is to investigate the approximation of P_w by hierarchical Markov and non-Markov models.

1. INTRODUCTION

Wavelet shrinkage is a widely-used and effective method for image modeling and enhancement problems. However, virtually all wavelet marginal models currently being used in wavelet shrinkage [7] assume the coefficients to be decorrelated and treated individually. Although such independent models result in simple nonlinear shrinkage algorithms, this approach is not optimal in the sense that for most spatial statistics or prior models the wavelet transform is not a perfect whitener.

There have been several recent approaches that examine the joint statistics of the coefficients. Xu et al. [9] used the scale dependent consistency between wavelet coefficients for the denoising process. In separate work by Simoncelli [8] and Crouse et al. [3], probabilistic models were studied that capture wavelet coefficient dependencies, mainly across scales. Crouse et al. [3] considered hidden states describing each coefficient's significance. Instead of the coefficients values, they propose statistical models for a coefficient's hidden state dependencies. 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.

Having been motivated by these inter-coefficient probabilistic studies, the primary goal of this work is to propose a well-structured wavelet-domain correlation model which is capable of describing coefficient dependency by introducing a local neighborhood containing statistics of within- and across-scale coefficients. The main novelty is the systematic approach we have taken to define a wavelet-based neighborhood system consisting of 1) inter-scale dependency evolution, 2) within-scale clustering, and 3) across-orientation (geometrical constraints) activities. This probabilistic modeling is directly applied to the coefficient values, but to some extent their significance is also considered.

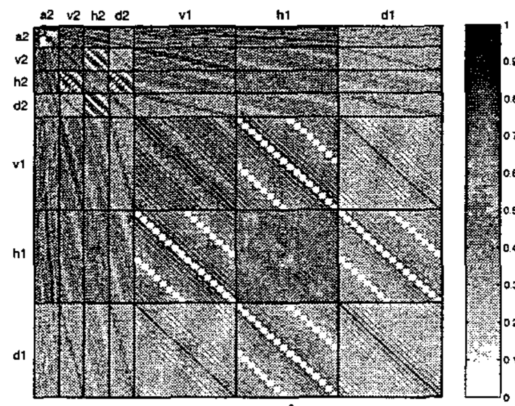

Fig. 1. Correlation coefficients of a spatial thin-plate model in the wavelet domain. The main diagonal blocks correspond to the same scale and orientation, whereas off-diagonal blocks illustrate cross-correlations across orientations or across scales.

It is well-known that the wavelet-domain covariance, P_w (Figure 1), is block-structured. We have observed [1] that, although the majority of correlations are very close to zero (i.e., decorrelated), a relatively significant percentage (10%) of the coefficients are strongly correlated across several scales or within a particular scale but across three orientation subbands. One approach to statistically model these relationships was to implement a multiscale model [1]. Although the MS model captured the existing strong parent-child correlation, spatial and inter-orientation interactions are not explicitly taken into consideration. Our most recent work [2] investigated the significance of inter-orientation

0-7803-7997-7/03/$17.00 © 2003 IEEE

and spatial relationships, which we seek to model more formally in this paper.

2. WAVELET COVARIANCE APPROXIMATION

Suppose we have a smooth Gaussian Makov Random Field (MRF) prior P_s, projected into the wavelet domain with wavelet prior P_w as is shown in Figure 1. Following our past work in wavelet statistics [1, 2] we propose to model the wavelet coefficients not as independent, but as governed by some local stochastic process. Since correlations are present within and across scales, clearly a random field model for wavelet coefficients will need to be explicitly hierarchical.

Based on the correlation map of Figure 1, six different symmetric neighborhood structures are chosen. For a coefficient w_i belonging to the wavelet coefficients set $W = \{W_h, W_v, W_d\}$ we define

$$p_k(i) = \{p^1(i), \dots, p^k(i)\}$$
$$c_k(i) = \{c^1(i), \dots, c^k(i)\}$$

$$s_{ud}(i): \begin{matrix} \bullet \\ \times \\ \bullet \end{matrix} \qquad s_1(i): \begin{matrix} & \bullet & \\ \bullet & \times & \bullet \\ & \bullet & \end{matrix} \qquad s_2(i): \begin{matrix} \bullet & \bullet & \bullet \\ \bullet & \times & \bullet \\ \bullet & \bullet & \bullet \end{matrix}$$

$$s^2_{ud}(i): \begin{matrix} \bullet \\ \bullet \\ \times \\ \bullet \\ \bullet \end{matrix} \qquad s_{lr}(i): \bullet \times \bullet \qquad s^2_{lr}(w): \bullet \bullet \times \bullet \bullet$$

where $p^\alpha(i)$ is the ancestor of w_i of α generations (scales), $c^\alpha(i)$ is the set of descendants of w_i of α generations (scales), and $s^\alpha_n(i)$ defines various sibling sets (same scale as w_i). This allows us to propose six neighborhood structures:

$$N_1(i) = \{p_1(i), c_1(i), s_1(i)\}$$
$$N_2(i) = \{p_1(i), c_1(i), s_2(i)\}$$
$$N_3(i) = \{p_2(i), c_2(i), s_1(i)\}$$
$$N_4(i) = \{p_2(i), c_2(i), s_2(i)\}$$

$$N_5(i) = \begin{cases} \{p_2(i), c_2(i), s_2(i), s_{lr}(s_{ud}(v(i))), s_{ud}(d(i))\}, & if \;\; w_i \in W_h \\ \{p_2(i), c_2(i), s_2(i), s_{lr}(s_{ud}(h(i))), s_{lr}(d(i))\}, & if \;\; w_i \in W_v \\ \{p_2(i), c_2(i), s_2(i), s_{lr}(v(i)), s_{ud}(h(i))\}, & if \;\; w_i \in W_d \end{cases}$$

$$N_6(i) = \begin{cases} \{p_2(i), c_2(i), s^2_{ud}(i), s_{lr}(s_{ud}(v(i))), s_{ud}(d(i))\}, & if \;\; w_i \in W_h \\ \{p_2(i), c_2(i), s^2_{lr}(i), s_{lr}(s_{ud}(h(i))), s_{lr}(d(i))\}, & if \;\; w_i \in W_v \\ \{p_2(i), c_2(i), s_1(i), s_{lr}(v(i)), s_{ud}(h(i))\}, & if \;\; w_i \in W_d \end{cases}$$

where operators d, v, and h return diagonal, vertical, and horizontal subband counterparts. With these hypothesized structures in place, the remainder of this article develops and tests two associated models.

2.1. Local Estimation

We begin with an explicitly local estimator, where only those measurements within the neighborhood are used. Thus, given the noisy measurements

$$y_i = w_i + v_i, \qquad v_i \sim \mathcal{N}(0, r^2) \tag{1}$$

we form a local estimation problem

$$\mathbf{y}_i = [y_i \quad \mathbf{y}(N(i))], \qquad \mathbf{w}_i = [w_i \quad \mathbf{w}(N(i))]$$

from which the standard estimator follows trivially

$$\hat{\mathbf{w}}_i = \Lambda_{\mathbf{w}_i, \mathbf{y}_i} \cdot \Lambda^{-1}_{\mathbf{y}_i} \cdot \mathbf{y}_i \tag{2}$$

and where we are only interested in

$$\hat{\mathbf{w}}_i(1) = E[w_i | \mathbf{y}_i]$$

2.2. MRF-Based Estimation

The second approach is to use a local model, for which P_w^{-1} is sparse; that is, a Markov random field.

As can easily be verified, however, in most cases the wavelet prior is not indeed, Markov, therefore an approximate Markov model needs to be estimated. Our local model is

$$w_i = \sum_{j \in N(i)} g_{i,j} w_j + \eta_i \tag{3}$$

Note that, in distinct contrast to the vast majority of planar MRF models in which g is stationary, the structure of the wavelet tree (asymmetry between parent and child, or between siblings) makes g rather nonstationary and considerably complicates model estimation [4, 5].

3. MODEL EVALUATION

We examine six different wavelet neighborhood structures. Clearly each choice of neighborhood will differ in its statistical accuracy. The six local and MRF-based results are compared with the null estimator

$$\hat{w}_i = y_i$$

and the pointwise estimator

$$\hat{w}_i = y_i \frac{\sigma^2_w}{\sigma^2_w + \sigma^2_v}$$

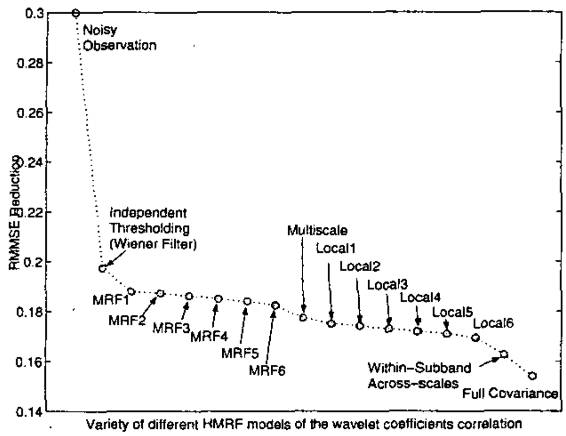

Fig. 2. RMMSE noise reduction as a function of HMRF neighborhood systems used in the two approximation techniques. MRF stands for MRF-based method and Local depicts explicit local method.

with the RMMSE of all cases plotted in Figure 2. It is clear that the vast bulk of the benefit is to be obtained from relatively few coefficients in the locality of the center coefficient. Empirically, the presence of within-scale (and across-orientation) correlation in these simulations (from $N_1(w)$ towards $N_6(w)$) reduces the estimation error.

The second aspect of comparison is computational complexity. The complexity is complicated by the presence of MRF models, which may be solved in a wide variety of ways. In increasing order of complexity we have a) Pointwise, b) Local 1-6, c) Multiscale, d) MRF 1-6, e) Full model.

Clearly the pointwise method is a linear approach, known as Wiener filtering, with its complexity growing linearly as the number of wavelet coefficients n increases. On the other side, the complexity of the Multiscale-based estimator is $O(d^3 n)$, where d shows dimensionality of every node (in the simplest case $d = 1$) [1]

Let us re-write (2) to investigate the complexity of local models

$$\hat{\mathbf{w}}_i = \Lambda_{\mathbf{w}_i,\mathbf{y}_i} \cdot \Lambda_{\mathbf{y}_i}^{-1} \cdot \mathbf{y}_i = L_i \mathbf{y}_i \qquad 1 \leq i \leq n \qquad (4)$$

with matrix L_i of size $m \times m$, where $m \ll n$ denotes the neighborhood size. Complexity of calculating L_i is of order $O(m^3)$. The prior model can be stationary or not. We consider both cases at this point:

- stationary prior model:

  In this case the complexity of the model estimation process L is fixed to $O(m^3)$, because $L_i = L_j$, if $j \neq i$. Thus total complexity of the estimation process $\hat{\mathbf{w}} = L\mathbf{y}$ is $O(m^3 + n \cdot m^2)$.

- prior model is not stationary:

  It is known that a stationary prior projected into the wavelet domain, changes to nonstationay, because of the multiscale nature of the wavelet domain. In this case the complexity of the model estimation process L is $O(n \cdot m^3)$, because $L_i \neq L_j$, if $j \neq i$. Thus, the total complexity of the local estimator $\mathbf{w} = L\mathbf{y}$ is $O(n \cdot (m^3 + m^2))$.

The computational cost for the MRF-based estimators is more complicated. In this paper, we only consider the simple linear case, i.e, Gaussian prior. Let us consider these two pieces of information, respectively the MRF prior (3) and the measurement:

$$\begin{aligned} G\mathbf{w} &= \eta \qquad & \eta &\sim \mathcal{N}(0, Q) \\ \mathbf{y} &= \mathbf{w} + \mathbf{v}, \qquad & \mathbf{v} &\sim \mathcal{N}(0, R) \end{aligned}$$

Define a linear estimator to find $\mathbf{w}$ which minimizes

$$\begin{aligned} &||\mathbf{y} - \hat{\mathbf{w}}||_{R^{-1}} + ||G\hat{\mathbf{w}}||_{Q^{-1}} \\ &\Rightarrow (R^{-1} + G^T Q^{-1} G)\hat{\mathbf{w}} = R^{-1}\mathbf{y} \end{aligned} \qquad (5)$$

which is a linear system of equations to be solved. We examined iterative solvers such as Guass-Sidel and PCG to solve for $\mathbf{w}$. The computational complexity of these algorithms is believed to be $O(itrn \times m \times n)$, where m is neighborhood size and $itrn$ shows number of iterations for a solver to converge to a predefined tolerance value. The experimental results indicate surprisingly fast convergence speed. The PCG algorithm was run for six different MRF-based neighborhood systems. In these experiments a thin-plate prior model with 3 different correlation lengths was considered. For the purpose of simplicity we examined textures of 32×32 size projected by Daubechies wavelets.

Figure 3 illustrates the $itrn$ number for the PCG to solve (5) for all six neighborhood systems, where the mother wavelet was fixed to be db2. A thin-plate prior was simulated with three different correlation lengths. The results indicate that for a fixed correlation length $itrn$ number remains almost unchanged. However, the increment of correlation length, i.e., larger extent of pixels connectivity (smoothness), increases the computational cost for the estimator.

To examine the sensitivity of the proposed MRF-based technique, various Daubechies wavelets were investigated. Figure 4 shows the $itrn$ number for the PCG for all $N1 - N6$ systems, where a thin-plate model with a fixed correlation length was used. The $itrn$ remains small for all cases and increases where more smooth wavelet is considered. It shall be noticed that in all experiments the $itrn$ is a relatively small number which represents the low complexity for our wavelet linear MRF-based estimator.

Figure 5 displays estimation error reduction based on each correlation model as a function of estimator's time

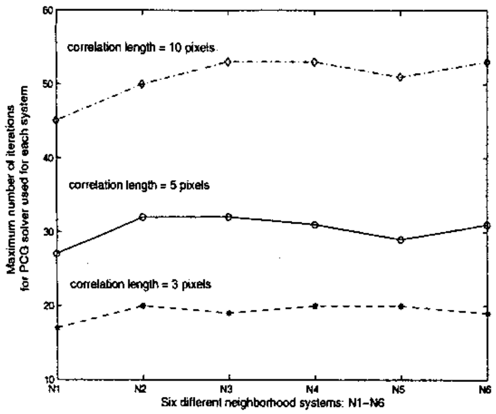

Fig. 3. Maximum number of iterations for the PCG solver to converge to a predefined tolerance as a function of wavelet domain MRF neighborhood systems proposed in this paper. These simulations were run for thin-plate prior texture with three different correlation lengths. The mother wavelet was considered to be db2.

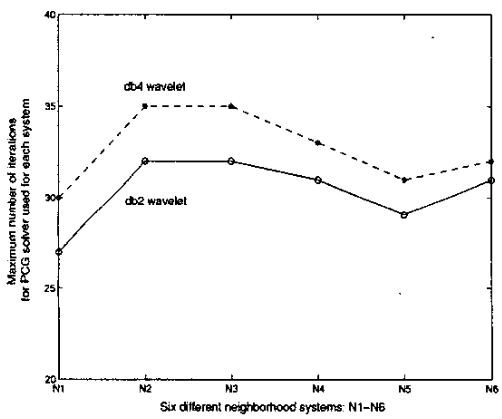

Fig. 4. Maximum number of iterations for the PCG solver to converge to a predefined tolerance as a function of wavelet domain MRF neighborhood systems proposed in this paper. The simulations were run for thin-plate prior texture with correlation length to be 5 pixels. The mother wavelets db2 and db4 were considered.

complexity. The local-based estimators achieve lower error rates, but with higher computational burden.

4. CONCLUSIONS

A thorough 2-D wavelet covariance study has been presented in this paper. An examination of the coefficient correlations, within or across scales, revealed the fact the there exist local stochastic models (explicit or MRF) governing these local dependencies. The proposed hierarchical random fields model exhibits a sparse neighborhood structure which absorbs correlation of the given coefficient with the

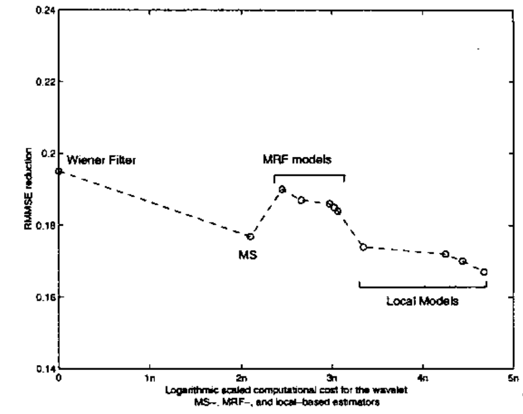

Fig. 5. RMMSE noise reduction as a logarithmic function of computational complexity associated with estimators based on the proposed wavelet correlation models. n show number of coefficients.

rest of the wavelet tree.

Following the modelling stage, the model accuracy was evaluated by comparing it with the common thresholding methods in the RMMSE sense and estimating its computational complexity. The principle motivation of this work is to devise an estimation or denoising algorithm which takes into account this probabilistic model and results in optimum error and low computational cost.

5. REFERENCES

[1] Z. Azimifar, P. Fieguth, and E. Jernigan. Hierarchical multiscale modeling of wavelet coefficients correlations. *Proceedings of the 9th SSPR*, 2002.

[2] Z. Azimifar, P. Fieguth, and E. Jernigan. Towards random field modeling of wavelet statistics. *Proceedings of the 9th ICIP*, 2002.

[3] M. Crouse, R. Nowak, and R. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE trans. on Signal Processing*, 46:886–902, 1998.

[4] R. Dufour and E. Miller. Statistical signal restoration with $1/f$ wavelet domain prior models. *Signal Processing*, 78:209–307, 1998.

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

[6] S. Lakshmanan and H. Derin. Gaussian markov random fields at multiple resolutions. *Chellappa and Jain editors: Markov Random Fields: Theory and Applications*, pages 131–157, 1993.

[7] D. Mumford and J. Huang. Statistics of natural images and models. *Proceedings of the CVPR'99*, 1999.

[8] E. Simoncelli. Modeling the joint statistics of images in the wavelet domain. *Proceedings of the SPIE*, 1999.

[9] Y. Xu, J. Weaver, D. Healy, and J. Lu. Wavelet transform domain filters: a spatially selective noise filtration technique. *IEEE trans. on Image Processing*, 3(6):747–758, 1994.