CORRELATED WAVELET SHRINKAGE: MODELS OF LOCAL RANDOM FIELDS ACROSS MULTIPLE RESOLUTIONS

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

Department of Systems Design Engineering
University of Waterloo, Waterloo, Ontario, Canada

ABSTRACT

This paper proposes a novel correlated shrinkage method based on wavelet joint statistics. Our objective is to demonstrate effectiveness of the wavelet correlation models [1] in estimating the original signal from a noising observation. Simulation results are given to show the advantage of the new correlated shrinkage function. In comparison with the popular nonlinear shrinkage algorithms, it improves the denoised results.

1. INTRODUCTION

Recent work in wavelet statistics had led to a growing realization that modeling wavelet coeffi cientsas independent, or at best correlated only across scales, may be a poor assumption. While recent developments in wavelet-domain Hidden Markov Models [2, 3] (notably HMT-3S [4]) account for within-scale dependencies, we fi nd empirically that wavelet spatial statistics are strongly orientation dependent [1], structures which are surprisingly not considered by state-of-the-art wavelet modeling techniques.

We studied structure of the existing wavelet correlations by: 1) fi ndingthe wavelet sample covariance over a large collection of real images, 2) adopting the standard 2-D wavelet transform diagram to display the locality among the coeffi cients,as is illustrated by Fig.1. Each panel is associated with one of the tree subbands, and illustrates the correlation map for a given coeffi cient(•) with its neighborhood across subbands and scales. Details of this statistical study, in addition to our development of the multiscale as well as Markov random fi eldmodels to describe the exhibited wavelet neighborhood structure can be found in [5, 1].

Given the above observation of wavelet joint statistics, this paper is focused on the development of a non-linear correlated empirical Bayesian shrinkage algorithm, with illustrations and evaluations of its estimation results.

2. INDEPENDENT WAVELET SHRINKAGE

Suppose a random fi eld $\underline{x}$ is projected into the wavelet domain with a resulting coeffi cientvector $\underline{w}$. The objective is

Supports of the NSERC and CITO of Canada are acknowledged.

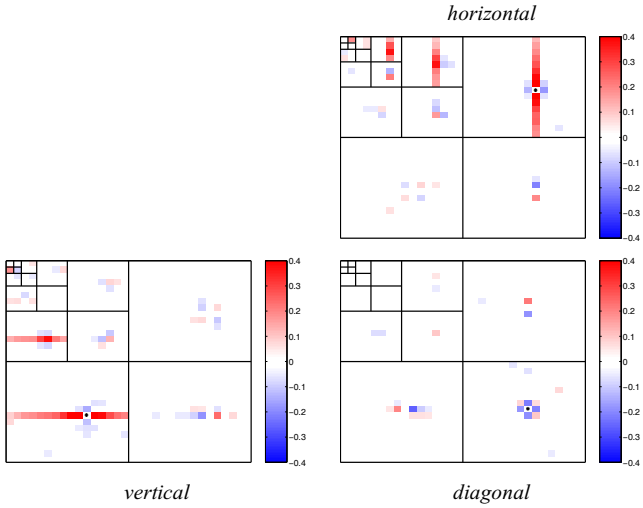


Fig. 1. Wavelet (db2) correlation structures averaged over a collection of 5000 real images. Each panel is associated with one of the tree subbands, and illustrates the correlation map for a given coeffi cient(•) with its local neighborhoods across subbands and scales.

to estimate $\underline{\hat{w}}$, given the noisy observation $\underline{y}$:

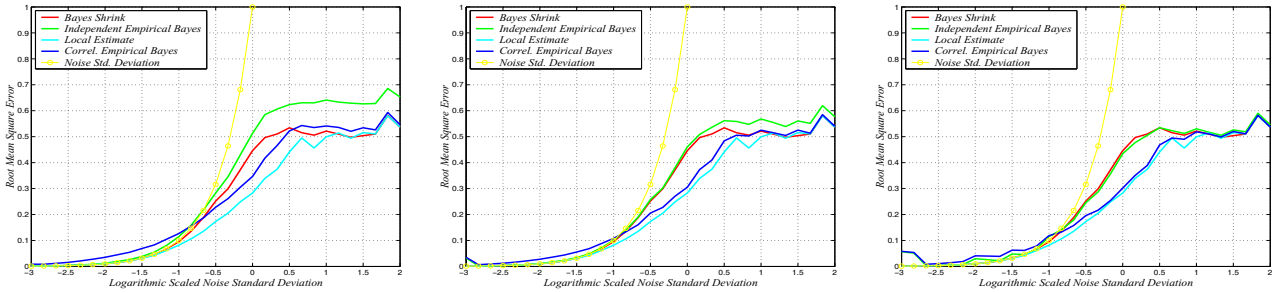
$$\begin{array}{rcll} \underline{y} & = & \underline{w} + \underline{\nu} & \underline{\nu} \sim \mathcal{N}(\underline{0}, \Sigma_\nu) \\ y_i & = & w_i + \nu_i & \nu_i \sim \mathcal{N}(0, \sigma_\nu^2) \end{array}$$

where $\underline{\nu}$ is assumed additive *i.i.d.* random noise. In general, if the coeffi cients are assumed *independent* and normally distributed, then the linear Bayesian estimate is optimum in mean squared error sense

$$\hat{w_i} = E[w_i|y_i] = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_\nu^2} y_i \qquad (1)$$

However, since the wavelet marginal prior is well-known to be non-Gaussian, then $E[\underline{w}|\underline{y}]$ is a non-linear process. One of the superior non-linear shrinkage methods, known as *BayesShrink* [6], determines threshold $T_{Bayes} = \frac{\sigma_\nu^2}{\sigma_w}$ for each subband assuming a Generalized Gaussian Distribution (GGD) for the coeffi cients. Chang *et al.* [6] observed that the threshold value T_{Bayes} is very close to the optimum threshold. *BayesShrink* performs soft thresholding, with its data-driven, subband dependent threshold. The results obtained by *BayesShrink* visually look more appealing than those obtained using *VISUShrink* and *SUREShrink*.

0-7803-9134-9/05/$20.00 ©2005 IEEE



(a) windowsize 2 (b) windowsize 6 (c) windowsize 30

Fig. 2. Plots of RMSE measurement for db2 BayesShrink as well as independent and correlated empirical Bayesian shrinkage, with thin-plate as the prior and uniform averaging window sizes of 2, 6 and 30.

All of these shrinkage algorithms treat the non-Gaussian coeffi cients as independent, however based on our observations of the wavelet joint statistics we propose a correlated shrinkage method whose non-linearity is approximated through an empirical Bayesian approach.

3. EMPIRICAL BAYESIAN ESTIMATION

The GGD prior for wavelets is, at best, a heuristic, or approximation. Different classes of images will necessarily have different wavelet priors. It is, therefore, very diffi cult to talk about or even formulate the optimum Bayesian estimates, making an empirical approach attractive.

Given a vast number of $\{w_i, y_i\}$ pairs, the optimum Bayesian expectation can be formulated as a sample mean

$$\hat{w}_i = E[w_i|y_i] \simeq mean\{w_j|y_j \simeq y_i\} \qquad (2)$$

is a non-linear shrinkage, while $\{w_i\}$ are independent.

To defi nethe joint Bayesian estimate, it must be noticed that

$$E[w_i|\underline{y}] \neq E[w_i|y_i]$$

because the $\{y_i\}$ are not assumed independent because of the correlation in the $\{w_i\}$ [1, 5].

To solve the joint estimate one normally limits the attention to some neighborhood $\mathcal{N}$

$$E[w_i|\underline{y}] \simeq E[w_i|\{y_j; j \in \mathcal{N}_i\}] \qquad (3)$$

In principle (3) can be solved as before, using empirical Bayes, but where we now take a sample mean over similar neighborhoods

$$E[w_i|\underline{y}] \simeq mean\{w_k|y_l \simeq y_j;\ \ l = \mathcal{N}_{k,m}\ \ j = \mathcal{N}_{i,m}\} \qquad (4)$$

where $\mathcal{N}_{i,m}$ is the m^{th} element index in the neighborhood of i. However, the required data grows exponentially with neighborhood size and is impractical for all but the smallest neighborhoods.

Instead, we can imagine combining (2) and (3), using a linear method to take into account joint relationship, and empirical Bayes to infer any needed non-linearity to fi nda good estimate. The development of such an approach follows next.

4. CORRELATED SHRINKAGE

Our premise is that the wavelet coeffi cients are correlated, thus a neighborhood structure must fi rstbe defi ned.

Based on the correlation map of Fig. 1, one can defi ne various different neighborhoods, of which only two structures are proposed here. For a coeffi cient w_i belonging to the wavelet coeffi cientsset $\underline{w} = \{\underline{w}_h, \underline{w}_v, \underline{w}_d\}$ we defi ne

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

$s_{ud}(i)$: $s_{lr}(i)$: $s_2(i)$:

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 $s_n(i)$ defi nesvarious sibling sets (on the same scale as w_i). This allows us to propose two asymmetric neighborhood structures:

$$\mathcal{N}_1(i) = \begin{cases} \{s_{ud}(i), p_1(i)\}; & w_i \in \underline{w}_h \\ \{s_{lr}(i), p_1(i)\}; & w_i \in \underline{w}_v \\ \{s_{lr}(i), s_{ud}(i), p_1(i)\}; & w_i \in \underline{w}_d \end{cases}$$

$$\mathcal{N}_2(i) = \begin{cases} \{s_{ud}(i), s_2(v(i)), s_{ud}(d(i)), p_1(i)\}; & w_i \in \underline{w}_h \\ \{s_{lr}(i), s_2(h(i)), s_{lr}(d(i)), p_1(i)\}; & w_i \in \underline{w}_v \\ \{s_2(i), s_{lr}(v(i)), s_{ud}(h(i)), p_1(i)\}; & w_i \in \underline{w}_d \end{cases}$$

where operators d, v, and h return diagonal, vertical, and horizontal subband counterparts. With these hypothesized structures in place, the remainder of this section develops correlated wavelet shrinkage:

1. The given random fi eld$\underline{x}$ is projected into the wavelet domain with the resulting coeffi cient vector $\underline{w}$. A neighborhood system $\mathcal{N}$ is chosen.

$$y_i = w_i + \nu_i$$

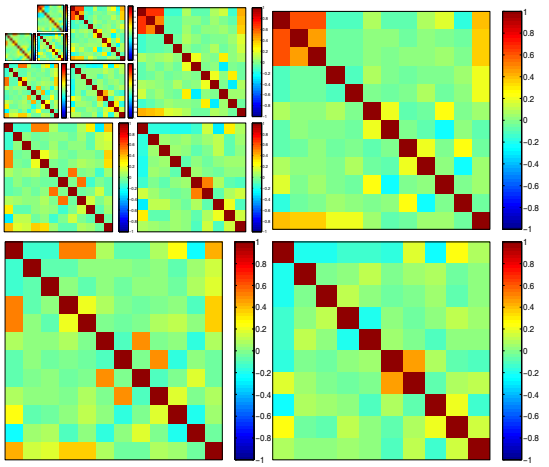


Fig. 3. The scale-dependent sample covariances we obtained for four-level wavelet transform of the "Goldhill" image. The order of the entities in each sample covariance is associated with that of the elements in $\mathcal{N}_2$.

Let us form two neighborhood vectors:

$$\begin{aligned} \underline{y}_i &= [y_i, \{y_j; j \in \mathcal{N}_i\}]^T \\ \underline{w}_i &= [w_i, \{w_j; j \in \mathcal{N}_i\}]^T \end{aligned}$$

2. If $\underline{w}_i$ is assumed jointly Gaussian (as an approximate assumption), an intermediate linear relaxing operation on the noisy coeffi cientsis

$$\underline{z}_i = P_{\underline{w}_i,\underline{y}_i} \cdot P_{\underline{y}_i}^{-1} \cdot \underline{y}_i \qquad (5)$$

where we are only interested in

$$z_i = \underline{z}_i(1) = E[w_i|\underline{y}_i]$$

For every individual wavelet coeffi cient the quantities $P_{\underline{w}_i,\underline{y}_i}$ and $P_{\underline{y}_i}$ are obtained numerically (by sampling).

3. The estimate $\hat{w}_i$ is found via empirical Bayes

$$\hat{w}_i = E[w_i|z_i] \simeq average\{w_j|z_j \simeq z_i\}$$

4. The computation of $mean\{\cdot\}$ can be done in different ways, such as uniform or triangular windowing.

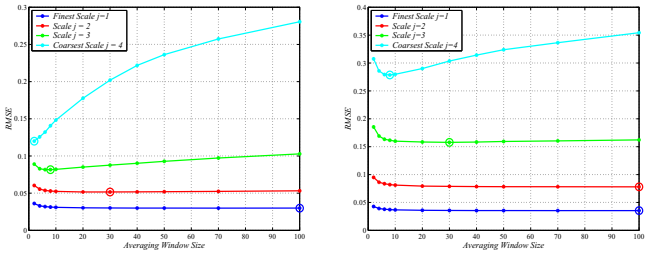
A schematic display of our Correlated empirical Bayesian Shrinkage (CBS) algorithm is

$$\underline{x} \xrightarrow{W} \underline{w} \xrightarrow{corrupted} \underline{y} \xrightarrow{local\ map} \underline{z} \xrightarrow{empirical\ Bayes} \underline{\hat{w}} \xrightarrow{W^{-1}} \underline{\hat{x}} \qquad (6)$$

5. EXPERIMENTAL RESULTS

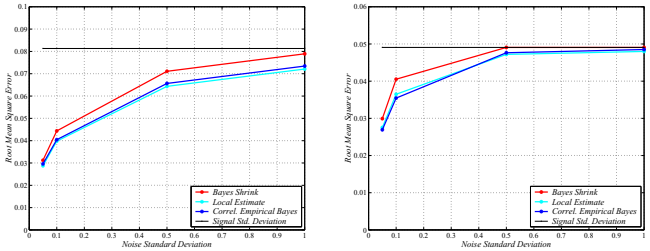
To test the performance of the proposed CBS algorithm, it was applied on a class of Gaussian Markov random fi elds, as well as real images and the simulation results were compared with that of *BayesShrink* and independent empirical Bayesian estimate (2).

CBS and Gauss Markov Random Fields: Sample statistics were found over a class of GMRF, including five textures (grass, pigskin, tree-bark, calf leather, and thin-plate). The averaged sample covariance over all five fi eldswas used to obtain $\underline{z}$ in (6) with association of the neighborhood structures defi ned in Sec. 4. The non-linear uniform averaging was then adopted to estimate $\underline{\hat{w}}$ in (6).



(a) $\sigma_\nu = 0.1$ (b) $\sigma_\nu = 0.5$

Fig. 4. RMSE of the CBS method calculated at several scales as a function of the averaging window sizes. The optimum window size at each scale (○) depends on the resolution as well as the additive noise level σ_ν.



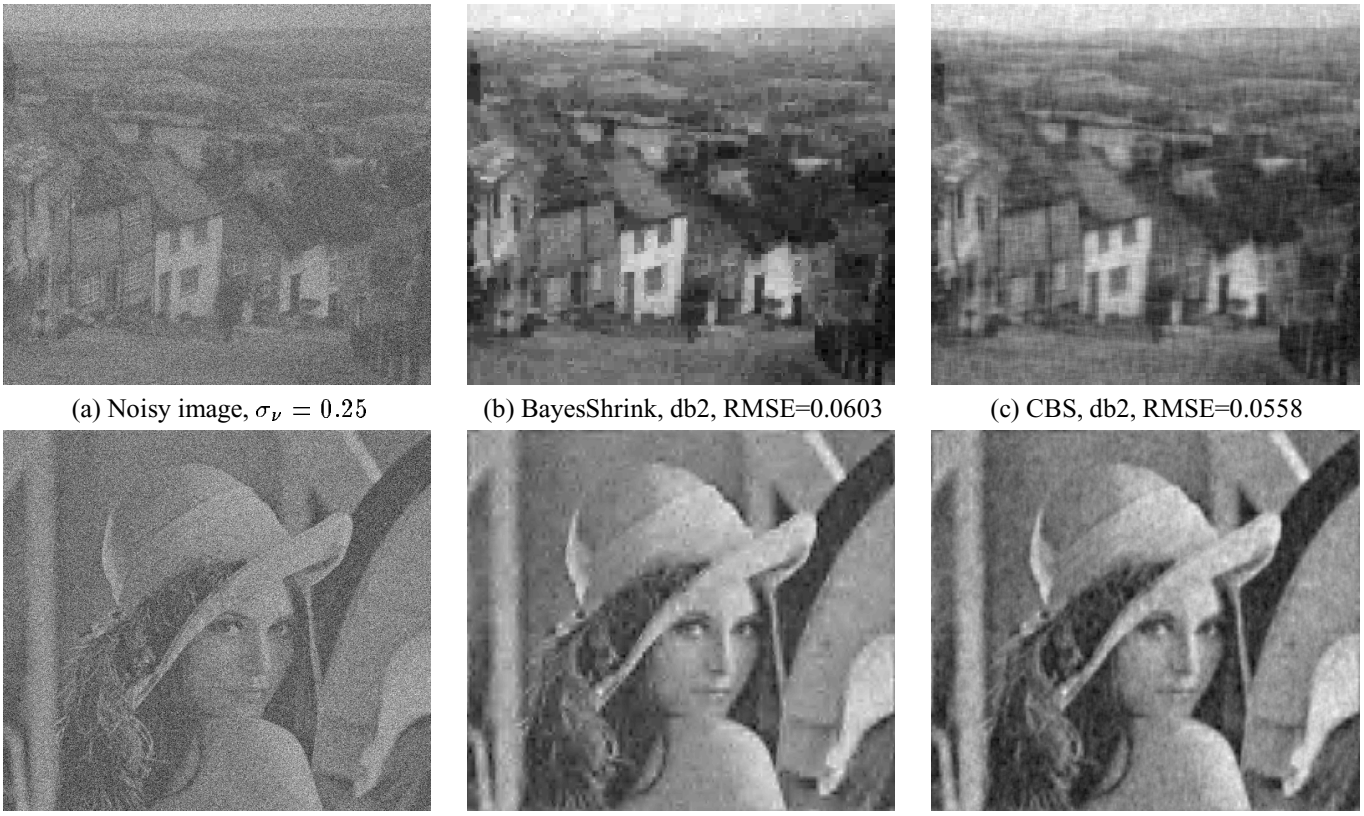
(a) $J = 4$ (b) $J = 2$

Fig. 5. RMSE comparison of *BayesShrink* and our CBS algorithm applied on a real image as a function of noise level and wavelet decomposition level J. The proposed CBS always results in lower estimation error.

Fig. 2 plots RMSE for BayesShrink as well as independent and correlated empirical Bayesian shrinkage with various averaging window size, applied on a GMRF corrupted with a large range of noise variances. As the window size gets bigger, superiority of the proposed CBS becomes evident. The results were obtained by using the neighborhood system $\mathcal{N}_2$.

CBS and Real Images: The above framework was also applied on several standard real images. For each test image, the local covariances, *i.e.*, $P_{\underline{w}_i}$ in (5), were calculated. Fig. 3 displays the scale-dependent sample covariances we obtained for four-level wavelet transform of the "Goldhill" image. In this experiment $\mathcal{N}_2$ was used, *i.e.*, the order of the entities in each sample covariance is associated with that of the elements in $\mathcal{N}_2$. The striking consistency between the plots in Fig. 1 and Fig. 3 is very interesting. These scale-dependent local maps were substituted in (5), completing the local estimation part.

The next subtlety was the notion of local averaging, which was studied in this simulation. As is illustrated by Fig. 4, the averaging window size depends on the resolution as well as the additive noise level. The coarser the resolu-



(a) Noisy image, $\sigma_\nu = 0.25$ (b) BayesShrink, db2, RMSE=0.0603 (c) CBS, db2, RMSE=0.0558

(d) Noisy image, $\sigma_\nu = 0.25$ (e) BayesShrink, db2, RMSE=0.0583 (f) CBS, db2, RMSE=0.0554

Fig. 6. The proposed CBS algorithm was successful to improve the artifacts appear in *BayesShrink* results and to depict more clear edges.

tion (*i.e.*, the less information we have), the smaller is the averaging window size.

Each panel in Fig 5 compares *BayesShrink* and our CBS algorithm applied on a real image in RMSE sense with different wavelet decomposition level J. It is evident that regardless of decomposition level, the CBS works better. The visualization by Fig. 6 shows the success of our CBS algorithm in removing the artifacts appear in *BayesShrink* results and in depicting more clear edges.

6. CONCLUSIONS

In this paper, we proposed a new shrinkage scheme with considerable improvement over the performance of the well-known shrinkage methods. Our CBS algorithm adopts joint statistics of the underlying image, resulting in a smaller estimation error and better visualization.

The ongoing research direction is to challenge the notion of non-Gaussianity for real images. The proposed CBS will be further investigated for a large class of real images with highly non-Guassian joint statistics and improvements of our model-based algorithm over other shrinkage methods with their GGD assumption for the coeffi cientsprior will be studied.

7. REFERENCES

[1] Z. Azimifar, P. Fieguth, and E. Jernigan, "Textures and wavelet-domain joint statistics," *Springer-Verlag: Lecture Notes in Computer Science: Image Analysis and Recognition*, vol. 3212, pp. 331–339, 2004.

[2] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE trans. SP, (46)886-902*, 1998.

[3] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree-structured imaged modeling using wavelet-domain hidden markov models," *IEEE trans. IP*, vol. 10, pp. 1056–1068, 2001.

[4] G. Fan and X. Xia, "Wavelet-based texture analysis and synthesis using hidden markov models," *IEEE Trans. on Circuits and Systems*, vol. 50, pp. 106–120, 2003.

[5] Z. Azimifar, P. Fieguth, and E. Jernigan, "Hierarchical markov models for wavelet-domain statistics," *Proceedings of IEEE-SSP Workshop*, 2003.

[6] S. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE trans. IP*, vol. 9, pp. 1522–31, 2000.