CORRELATED NON-LINEAR WAVELET SHRINKAGE

M. Amiri[1] *Z. Azimifar*[1] *P. Fieguth*[2]

[1]Department of Computer Science and Engineering, Shiraz University, Shiraz, Iran
amiri@cse.shirazu.ac.ir, azimifar@cse.shirazu.ac.ir
[2]Systems Design Engineering, University of Waterloo, Ontario, Canada
pfieguth@uwaterloo.ca

ABSTRACT

This paper examines non-linear shrinkage methods specifically taking into account the correlation structure of the multiresolution wavelet coefficients. In contrast to hidden Markov trees, which model the relationship of wavelet variance from scale to scale, here we wish to take advantage of coefficient correlation. A linear shrinkage based on the LLS (Linear Least Square) estimator, employing a sample correlation scheme, is tested and verified to have an aesthetic denoising performance. Then, state-of-the-art independent shrinkage functions are applied to exploit the efficiency of such techniques and to introduce non-linearity into the algorithm to compensate for non-Gaussianity of the wavelet statistics. The performance of the non-linear shrinkage technique, as used individually and together with the linear correlated approach, are illustrated.

***Index Terms*—** Wavelet joint statistics, non-linear shrinkage

1. INTRODUCTION

A majority of works on wavelet analysis assume the wavelet transform to be a perfect whitener. That is, the transformed coefficients are independent. Recent studies however, recognize some correlation and prove to improve performance. While these studies use some heuristics to choose a correlation structure, we use empirical analysis of both real and synthetic images to elicit the real structure. An illustration of correlations seen in three wavelet coefficients (∗) in three subbands with other coefficients is presented in Fig. 1. The correlations are averaged over 5000 real images so, the resulting correlation map gives a good measure whether there exists correlation among wavelet transform coefficients of real images. We observed that not only is there strong linear dependence among these coefficients, but there is also some structure in the dependencies [1]. Beside witnessing the familiar intra-scale correlations (persistence property), we observed other forms of dependencies. For example, vertical subband coefficients tend to covary with their horizontal neighbors and horizontal subband coefficients have dependency with their vertical neighbors. In brief, there is much correlation across scales, significant correlation within scale and less correlation between orientations and the correlation structure is orientation dependent [1].

Therefore, an overall assumption for the correlations will fail to fit the real structure. While hidden Markov trees model the relationship of coefficient variances (as being at either low or high state), we model the actual relationship of the coefficient itself with other wavelet coefficients. The novelty here, is that we use the correlation structure testified in this correlation map to design neighborhood systems accommodating these relationships. Although larger neighborhood systems may better fit the correlation structure, smaller ones provide much faster algorithms. Thus, several varied size neighborhood systems are implemented to address this tradeoff (Fig. 2).

In our previous work [2] we showed the virtue of a linear approach, Local Estimate. Although its performance is quite satisfactory (even better than the acclaimed non-linear BayesShrink), it does not conform to the fact that the marginal distribution of a wavelet coefficient is heavy-tailed and is best described by the mixture of a low-variance noise distribution and a high-variance signal distribution which is a consequence of the compaction property of wavelet coefficients. Local Estimate simply takes the marginal distribution to be Gaussian (hence a linear estimate). Therefore, to combine the efficacy of Local Estimate in considering correlations and the advantage of non-linear methods in fitting the problem nature, we applied non-linear methods to the output of the estimation from Local Estimate.

To justify that we are making non-linear methods, plots of denoised coefficients versus noisy ones (describing shrinkage functions) are presented.

Given the above observation of wavelet joint statistics, this paper is focused on the development of non-linear correlated shrinkage algorithms, with illustrations and evaluations of their estimation results.

2. INDEPENDENT WAVELET SHRINKAGE

Suppose a random field $\underline{x}$ is projected into the wavelet domain with a resulting coefficient vector $\underline{w}$. The objective is to

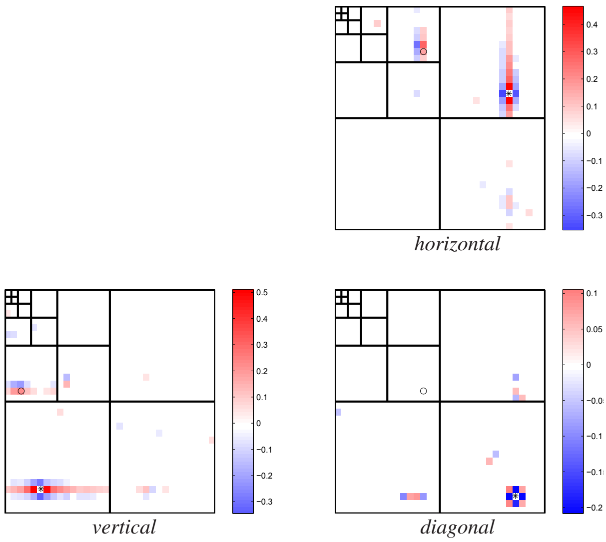

Fig. 1. Wavelet (db2) correlation structures averaged over a collection of 5000 real images. Each panel is associated with one of the tree subbands, and illustrates the correlation map for a given coefficient (∗) with its local neighborhoods across subbands and scales [2].

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 coefficients are assumed *independent* and normally distributed, then the linear Bayesian estimate is optimum in mean squared error sense

$$\hat{w}_i = E[w_i|y_i] = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_\nu^2} y_i \tag{1}$$

However, since the wavelet marginal prior is well-known to be non-Gaussian [4], then $E[\underline{w}|\underline{y}]$ is a non-linear process.

2.1. Empirical Bayesian Estimation

The GGD prior for wavelets is, at best, a heuristic, or an approximation. Different classes of images will necessarily have different wavelet priors. It is, therefore, very difficult 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

$$\hat{w}_i = E[w_i|y_i] \simeq mean\{w_j|y_j \simeq y_i\}. \tag{2}$$

This is a non-linear shrinkage, while no dependency among $\{w_i\}$ are taken into account.

To define the 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 due to the evidence of correlation in the $\{w_i\}$ [1].

To solve the joint estimate we need to consider multiple $\{y_i\}$ in some neighborhood $\mathcal{N}$

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

where $\mathcal{N}$ needs to be selected appropriate to the wavelet relationships, as illustrated in Fig. 1. In principle the joint estimate (3) can be solved using an empirical Bayes approach, as in (2), but we now compute a sample mean over similar neighborhoods

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

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 the joint relationships, followed by empirical Bayes inferring any needed non-linearity to find a good estimate. The development of such an approach follows in section 3.

2.2. SUREShrink

To overcome the problems of universal thresholding, adaptive denoising based on minimizing Steins Unbiased Risk Estimator (SUREShrink) was proposed [3]. SUREShrink is a scale dependent thresholding scheme which combines the universal threshold method with a scale-dependent adaptive selecting scheme. This method estimates the loss $E[(\hat{w}_i - w_i)^2]$ in an unbiased fashion:

$$SURE(\lambda; \underline{y}) = \underline{y} - 2|\{i : |y_i < \lambda|\}| + \sum_{i=1}^{d} min(|y_i|, \lambda)^2 \tag{5}$$

where $|.|$ shows the number of elements in a set.
For an observed vector $\underline{y}$ (the set of noisy wavelet coefficients in a subband), find the threshold λ_{SURE} that minimizes $SURE(\lambda; \underline{y})$, i.e., $\lambda_{SURE} = argmin_\lambda SURE(\lambda; \underline{y})$. The above optimization problem is computationally straightforward. This technique performs different global operations across scales. However, no spatial adaptation is assumed within each scale or each orientation.

2.3. BayesShrink

One of the superior non-linear shrinkage methods, known as *BayesShrink* [4], determines threshold $T_{Bayes} = \frac{\sigma_\nu^2}{\sigma_w}$ for each subband assuming a Generalized Gaussian Distribution (GGD) for the coefficients. Chang *et al.* [4] 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.

2.4. ProbShrink

Probshrink as another independent approach, takes the same prior for coefficients as BayesShrink (GGD) and estimates the probability of presence of signal of interest given an observed y_i to approximate the MMSE estimate of $\hat{w}_i$. Particularly,

$$\begin{aligned} \hat{w}_i &= E[w_i|y_i] \\ &\simeq P(H_1|y_i)y_i + 0 \\ &= \frac{f(y_i|H_1)P(H_1)}{f(y_i|H_1)P(H_1) + f(y_i|H_0)P(H_0)}\, y_i \end{aligned}$$

The approximation in the above formula is due to the assumptions that $E[w_i|H_1, y_i] = y_i$ and $E[w_i|H_0, y_i] = 0$. The conditional density of y_i is simply calculated from the convolution of the prior distribution of noise-free coefficients with the prior of noise (assumed zero mean Gaussian) which is the product of assuming noise is independent of signal. Parameters of the real signal prior are numerically calculated from the estimation of noise standard deviation, observation coefficients' standard deviation and observation coefficients' 4^{th} moment in each subband.

Evidently, independent wavelet shrinkage is still at the focus of shrinkage developments. This paper aims to exploit achievements of these independent shrinkages to boost up the advantage of correlated wavelet shrinkage. The correlated linear phase is achieved in (6).

All of these shrinkage algorithms treat the non-Gaussian coefficients 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. CORRELATED SHRINKAGE

To those familiar with the field of wavelet shrinkage, there are two main disciplines in correlated shrinkage including tree-based [5] and covariance-based methods [2, 6]. We follow the covariance-based scheme.

Our premise is that the wavelet coefficients are correlated, thus a neighborhood structure must first be defined.

Based on the correlation map of Fig. 1, one can define various different neighborhoods. A typical neighborhood $\mathcal{N}_2$ as defined in [2] is used for the experiments below. We now develop correlated wavelet shrinkage:

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

$$y_i = w_i + \nu_i$$

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 coefficients is

$$\underline{z}_i = P_{\underline{w}_i, \underline{y}_i} \cdot P_{\underline{y}_i}^{-1} \cdot \underline{y}_i \qquad (6)$$

where we are only interested in

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

For every individual wavelet coefficient the quantities $P_{\underline{w}_i, \underline{y}_i}$ and $P_{\underline{y}_i}$ are obtained numerically (by sampling). Thus far we derived the Local Estimate output.

3. The final estimate $\hat{w}_i$ is found via some non-linear shrinkage method. For the case of ProbShrink,

$$\hat{w}_i = E[w_i|z_i] \simeq P(H_1|y_i)y_i$$

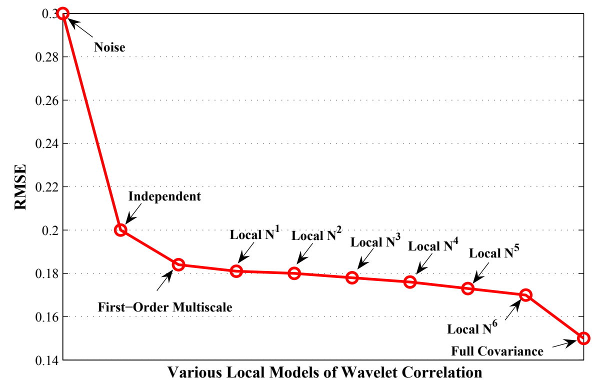

Fig. 2. Performance of an independent and some correlation-based wavelet denoising methods in terms of RMSE. The results are obtained on fair conditions where coefficient distributions are considered Gaussian. As is evident, all covariance-based shrinkage approaches outperform the independent approach. As the neighborhood is extended, more correlation is taken into account and better estimation is achieved. In the extreme case, full covariance gives the best achievable performance in the cost of high complexity [1].

A schematic display of our Correlated Shrinkage algorithm is

$$\begin{array}{ccccc} \underline{x} & \xrightarrow{\mathcal{W}} & \underline{w} & \xrightarrow{Corrupted} & \underline{y} \\ & & & & \downarrow Local-map \\ \hat{\underline{x}} & \xleftarrow[\mathcal{W}^{-1}]{} & \hat{\underline{w}} & \xleftarrow[Independent Shrinkage]{State-of-the-art} & \underline{z} \end{array}$$

4. EXPERIMENTAL RESULTS

The advantage of Local Estimate is verified on synthesized GMRF images [2]. Here, we apply our proposed combinational algorithms to a real image, Goldhill. The estimation performance of the algorithms as a function of noise strength is visualized in Fig. 3. As can be seen, all of the independent approaches are significantly improved when used together with a local estimator based on wavelet correlations. In particular, the Correlated Empirical Bayes is more efficient than the Independent Empirical Bayes. The performance of SUREShrink and ProbShrink are also significantly improved.

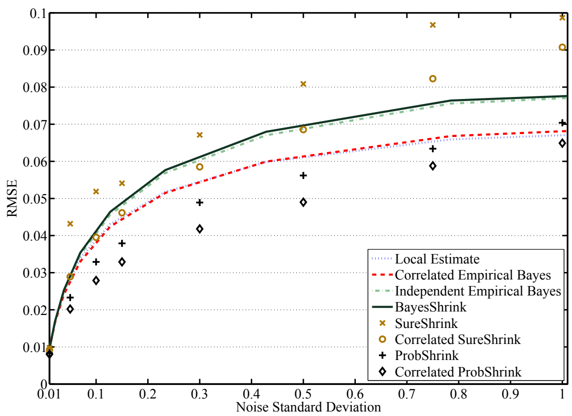

Fig. 3. Comparison of performance of different algorithms in different noise σ in terms of RMSE. As expected, the combined Empirical Bayes, SUREShrink and ProbShrink have improved performances over the corresponding original methods.

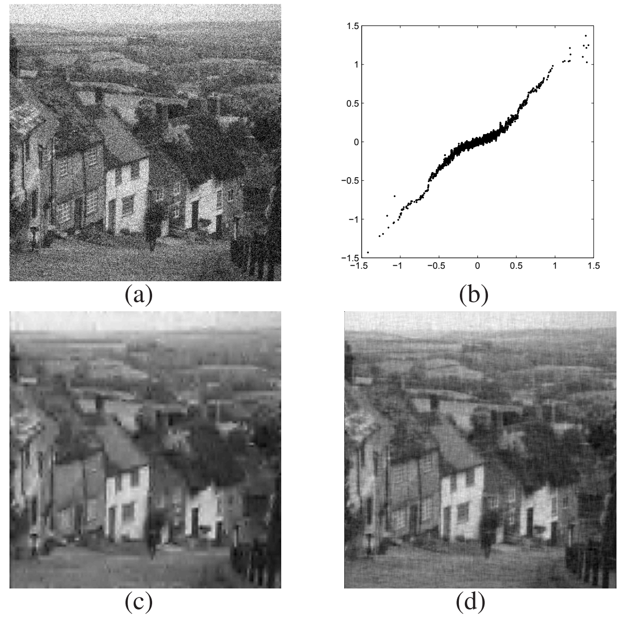

Fig. 4. Experimental Results: (a) Noisy image ($\sigma = 0.15$), (b) Non-linear transform function of Empirical Bayes, (c) Linear Local Estimate denoised image, (d) Combined Local Estimate and ProbShrink denoised image.

Additionally, considering the best performing hybrid technique (Correlated ProbShrink) as an example, visual quality of the denoised image is better than that of sole ProbShrink (Fig. 3 and Fig. 4). There are evident blurring artifacts in the ProbShrink image. Furtheremore, the mother wavelet (Daubechies 2) reveals itself in the image because of some sharp changes in the estimated wavelet coefficients due to the independence assumption. Our proposed method's denoised image has yet another desirable aspect, namely better edge preservation. Even fine edges as the window frames are preserved in the new method where as the individual independent approach fails to feature these singularities. Fig. 4(b) depicts the fact that the proposed approach, *i*.e., the combination of Local Estimate in considering correlations and the non-linear methods in fitting the problem nature, efficiently exhibits a non-linear shrinkage performance.

5. CONCLUSIONS

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

With this observations in place, the advantage of shrinkage algorithms taking advantage of wavelet correlations have been verified. In particular, there are some striking improvements in Fig. 3 which merit further study. An interesting research direction is to incorporate the correlation structures of Fig. 1into promising HMT's [5, 7, 8]. We shall use the observed relations to adapt an efficient tree model (or graph model as there are cyclic dependencies) for the correlations between hidden states instead of the coefficients themselves.

6. REFERENCES

[1] Z. Azimifar, "Image models for wavelet domain statistics," *PhD Thesis, Systems Design Engineering, University of Waterloo, Canada*, 2005.

[2] Z. Azimifar, P. Fieguth, and E. Jernigan, "Correlated wavelet shrinkage: Models of local random fields across multiple resolutions," *IEEE ICIP, Italy*, 2005.

[3] X. Zhang and M. Desai, "Nonlinear adaptive noise suppression based on wavelet transform," *IEEE Signal Processing Letters*, 1997.

[4] 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.

[5] 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.

[6] A. Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. on IP*, vol. 15, no. 3, pp. 654–665, 2006.

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

[8] 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.