J Math Imaging Vis
DOI 10.1007/s10851-011-0280-4

Empirical Study of Wavelet Domain Image Joint Statistics and Proposition of an Efficient Correlation Map

Zohreh Azimifar · Mehdi Amiri · Paul Fieguth · Ed Jernigan

Abstract This paper presents an empirical study of joint wavelet statistics for textures and other imagery to find an efficient correlation neighborhood. Since there is an established realization that modeling wavelet and other x-let coefficient relationships is crucial to any successful transform domain algorithm (such as Hidden Markov Trees), new works have been devoted to examine these dependencies from different aspects and propose an appropriate model. Because the time and computation complexity involved both in analyzing non-linear dependencies and in solving dependent models may restrict us to consider only a very small subset of contributing neighbors we focus our attention on linear dependencies (correlations) while having a squint on non-linear relations too. In this process, we study a collection of 5000 real images to corroborate our statistical analysis of the joint coefficient behavior and try to find an efficient and at the same time frugal relation map through different statistical means. The statistical observations are then certified by a coefficient significance measure and the competitiveness of the map is substantiated by plugging it into two dependent denoising frameworks.

Z. Azimifar (✉) · M. Amiri
School of Electrical and Computer Engineering,
Shiraz University, Shiraz, Iran
e-mail: azimifar@cse.shirazu.ac.ir

M. Amiri
e-mail: amiri@cse.shirazu.ac.ir

P. Fieguth · E. Jernigan
Systems Design Engineering, University of Waterloo, Waterloo, Canada

P. Fieguth
e-mail: pfieguth@uwaterloo.ca

E. Jernigan
e-mail: jernigan@uwaterloo.ca

Published online: 20 April 2011

Keywords Image modeling · Statistical dependency · Wavelets · Hidden Markov Tree

1 Introduction

Statistical models for joint pixel statistics are of central importance in many image processing applications. However long-range spatial interactions and the high dimensionality stemming from large problem size make statistical image modeling particularly challenging. Commonly this modeling is simplified by a change of basis, particularly using a wavelet transform (WT) [1]. Indeed, the WT has widely been used as an approximate whitener of statistical time series. It has, however, long been recognized [2, 3] that the wavelet coefficients are neither Gaussian, in terms of the marginal statistics, nor white, in terms of the joint statistics.

Wavelet marginal statistics have received considerable scrutiny. In general, a majority of the wavelet coefficients are small, with a few coefficients large in magnitude, making the marginal distribution more heavily tailed than a Gaussian, with a large peak near zero. Authors of [1] and [2] showed that this heavy tailed non-Gaussian marginal can be well approximated by a Gaussian Mixture Model (GMM), for example, a mixture of two normals or of one normal and one point mass function at zero [4, 5]. Bayesian estimation has been achieved with these non-Gaussian priors by using wavelet non-linear shrinkage [2], which takes into account this kurtotic behavior of the wavelet coefficients.

As opposed to marginal models, the question of joint models is much more complicated and admits far more possibilities, with statistical structures possible across different sub-bands, orientations, and scales. Since [6] proposed zero-tree coding for image compression there have been

many efforts to model joint structures using Markov random fields (MRFs) [7, 8], Besov spaces [9], wavelet hidden Markov models (HMMs) [2, 10, 11] and Gaussian scale mixtures (GSMs) [12]. Wavelet-based HMMs, in particular, have been thoroughly studied. These models can successfully outperform many other wavelet-based techniques in Bayesian denoising, estimation, texture analysis, synthesis, and segmentation.

Although a variety of joint models have been proposed and tested, few models appear to be obtained directly based on empirical studies of wavelet coefficient cross-statistics. Rather, they are based on intuitive or heuristic notions of wavelet neighborhood structures. Without an examination of the underlying statistics, such heuristic approaches necessarily leave unanswered questions of neighborhood sufficiency and necessity: are there important relationships remaining which are not being modeled, or are we adding unnecessarily to computational complexity by including irrelevant relationships?

Similar works have already been presented which address joint statistic analysis and modeling in x-let domains. Example of such works are [13–15] in Curvelet domain, [16] for Contourlets and [17] for wavelets. However, our work has its own characteristics. For instance, we have a linear approach as opposed to the works in [13, 14, 16, 17]. The reason is the complexity of involving extensive non-linear maps. Meanwhile, a linear map will be feasible enough to be able to accommodate larger maps. Moreover, we examine numerous candidate neighbors to decide if they correlated and if correlated whether they are beneficial to include. This is somehow similar to the work in reference [15] but different with the others which use a heuristically decided neighborhood. As another difference, we don't intend here to statistically model wavelet coefficients. Rather, we seek a correlation map that captures crucial linear dependencies and is efficient when used in related algorithms. This paper has unique features too. It tests the significance, usability, of individual neighbors to verify if they actually contribute to a better estimation of the reference coefficient. Also, the goal of this article is to definitively establish an efficient map. Therefore, the observations are made based on the wavelet coefficients of 5000 real images to keep generality.

Put briefly, the purpose of this paper is a systematic study of the joint linear correlations of wavelet coefficients including dependencies across scale, space, and orientation. We shall see that there are structures present in these statistics over a surprisingly large range of natural images and random fields. Clearly linear correlations model and represent only a subset of image structure, thus this paper is only one step in the broader goal (Fig. 1) of building more capable joint models for image representation.

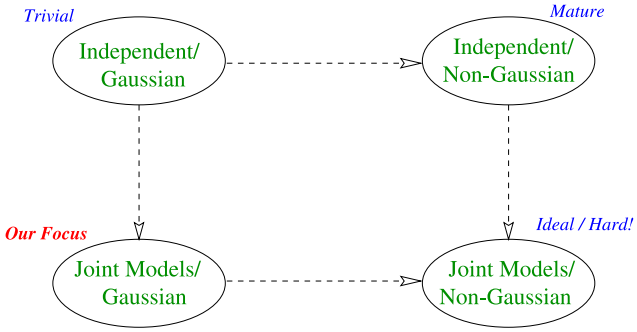


Fig. 1 Focus of this work: The development of joint Gaussian wavelet coefficient models with inter-coefficient dependencies

2 Wavelet Statistical Models

Researchers have proposed a variety of wavelet dependency models on the basis of the observed characteristics of wavelet coefficients: across-scale persistence, within-scale clustering [11], and sparse representation. In general, these models are divided into three main categories:

1. inter-scale [2, 11, 18],
2. intra-scale (spatial) [19–22], and
3. combined intra- and inter-scale [5, 7, 8, 10, 23, 24].

Our past works also focused on combined models, a wavelet multiscale statistical model [25], which captured parent-child correlations but not inter-orientation, and an approach to Markov modeling [26] across scales, orientations, and space. As a prevalent statistical framework for modeling wavelet coefficient distribution, HMMs have been adapted to accommodate all the three main categories. Therefore, we pursue examination of these categories with HMM terminology.

Since the introduction of Hidden Markov Models to wavelet joint statistics [2], significant work has been done to extend and improve the performance of these models. In general, these models adopt a probabilistic graph in which a wavelet coefficient w_i is associated with a discrete hidden state $s_i \in \{0, 1, \ldots, M-1\}$, thus modeling w_i as an M-state Gaussian mixture, conditionally independent of all of the other coefficients:

$$p(w_i, w_j|s_i) = p(w_j|s_i)p(w_i|s_i) \quad \forall i \neq j \tag{1}$$

A binary state, $M = 2$, is particularly common and is used to specify a low/high variance of w_i. Defining which pairs of hidden states can be conditionally independent and what the interpretation of the conditioning variable, several wavelet HMMs have been introduced.

2.1 Independent Mixture Model (IMM)

The simplest HMM, the IMM [2] (Fig. 2(a)), models the hidden states as independent, motivated by the observation

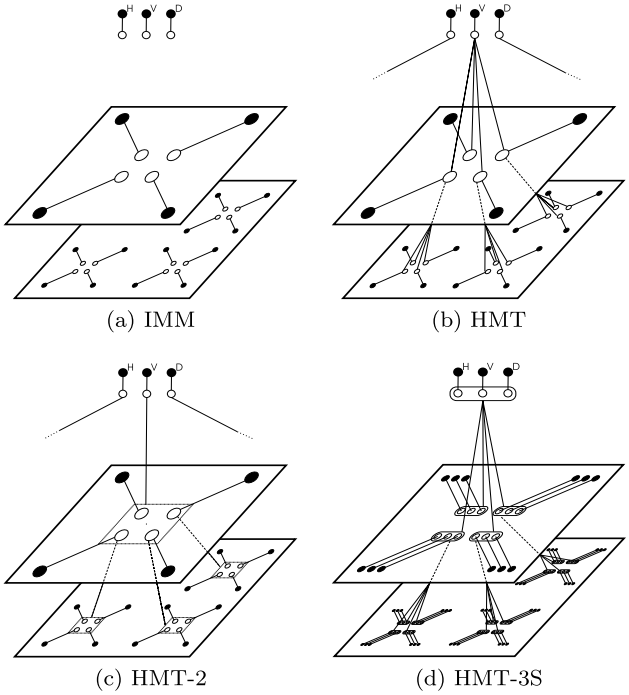


Fig. 2 Illustration of hidden Markov models. *Empty circles* denote hidden states and *filled circles* the coefficient values. (**a**) Independent hidden states; (**b**, **c**) Interscale dependencies; (**d**) Three subbands (H, V, D) integrated into one hybrid HMT

that the WT is an approximate whitening process, making the coefficients nearly decorrelated.

Although the IMM can well describe the wavelet marginal statistics as a Gaussian mixture, it does not attempt to characterize any wavelet-domain joint statistics. More advanced models, in which the hidden states have a Markovian dependency structure, are discussed below.

2.2 Hidden Markov Tree (HMT)

Crouse et al. [2] observed the persistence and clustering properties of wavelet coefficients and introduced the HMT (Fig. 2(b)) which captures inter-scale dependencies by imposing a tree structure on the hidden states across scales, while assuming independence within and across the three subbands. The statistics of the hidden state s_i are a function of the parent, $s_{\rho(i)}$, based on a transition probability

$$p_{s_i|s_{\rho(i)}}(s_i = m|s_{\rho(i)} = n) \tag{2}$$

It was shown in [11] that the HMT algorithm outperforms traditional wavelet-based techniques.

A generalization, to capture additional correlations between scales, is the HMT-2 [19], where the state s_i depends on the state of its parent $s_{\rho(i)}$, as before, but also on siblings of its parent (Fig. 2(c)) leading to higher-order hidden states. The approach is motivated by the correlation of the wavelet bases in two adjacent scales and the long length of the filters used in the decomposition process. HMT-2 empirical results show some improvement in signal denoising [19].

2.3 Contextual Hidden Markov Model (CHMM)

The HMT model focused on the inter-scale dependencies by imposing a tree structure in the wavelet domain. To support additional connectivity, the CHMM was developed by [2], which adds a context structure to model both inter-scale and intra-scale dependencies. The basic idea of the CHMM is to define contexts as a function of the wavelet coefficient w_j and its local neighbors to capture the spatial dependencies such that given the context the coefficients are treated as independent. The CHMM has many potential advantages [2] over traditional HMMs in exploiting the wavelet correlation structure, offering similar denoising performance with reduced computational complexity compared with the HMT.

The lack of spatial adaptability of the CHMM [19, 20] may limit its advantages in image processing tasks, so a Local CHMM was proposed [19], exploiting both interscale and local statistics.

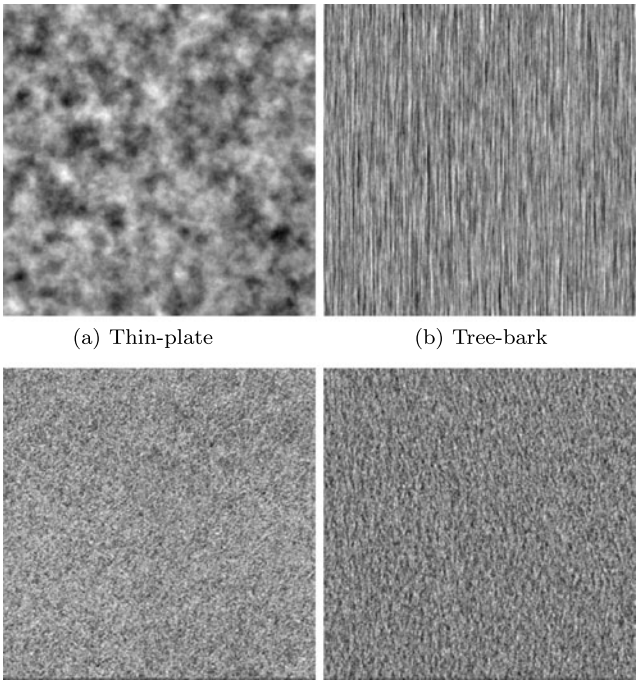
2.4 Hidden Markov Model-Three Subbands (HMM-3S)

The above joint models assume that the subbands of different orientations are independent, an almost universal assumption. Although this simplifies wavelet image modeling, for natural textures regular spatial patterns may result in noticeable dependencies across subbands [10]. The HMT-3S (Fig. 2(d)) includes the joint interscale statistics captured by the HMT, but adds dependencies across subbands by treating jointly the three hidden elements across the three orientations. The HMT-3S was successfully applied in texture analysis and synthesis with improved performance over the HMT models.

It is clear that a significant body of research addresses wavelet joint statistics and modeling. The more general the model, the broader the range of included statistics, the better the results. In this regard, only few models appear to be obtained directly based on empirical studies of wavelet coefficient cross-statistics, but more heuristic and intuitive. In the absence of an examination of the underlying joint statistics, such heuristic approaches necessarily leave unanswered questions of neighborhood sufficiency and necessity as stated above. Therefore, the following section will examine empirical wavelet correlations for the purpose of assessing model necessity and sufficiency.

3 Wavelet Neighborhood Modeling

Let us not face the problem in full details ab initio. To have a firm ground we start inspecting structured images with



(a) Thin-plate (b) Tree-bark

(c) Grass (d) Calf leather

Fig. 3 Four textures were used to generate wavelet statistics; random images from the corresponding GMRFs are shown

known pixel interaction models, synthetic statistical textures. As wished for, their correlation matrix, can be exactly derived from their kernels and their neighborhood structure easily identified.

Consider a statistical texture $\mathbf{x}$ with known covariance $P_{\mathbf{x}}$. The fine scale texture $\mathbf{x} \equiv \mathbf{x}^0$ has a 2-D wavelet decomposition

$$\mathcal{W}\mathbf{x}^0 = \{\mathbf{w}^1, \mathbf{w}^2, \ldots, \mathbf{w}^J, \mathbf{x}^J\} \tag{3}$$

where $\mathbf{w}^j = \{\mathbf{w}_h^j, \mathbf{w}_v^j, \mathbf{w}_d^j\}$ contains the three orientation subbands at scale j, and where $\mathbf{x}^J$ represents the scaling coefficients at the coarsest scale J. Because it is so widely used in wavelet HMMs, we restrict our attention to the Daubechies class of wavelets.

The wavelet operator $\mathcal{W}$ is linear, therefore the known spatial domain covariance structure $P_{\mathbf{x}}$ can be projected into the wavelet domain

$$P_{\mathbf{w}} = \mathcal{W} P_{\mathbf{x}} \mathcal{W}^T. \tag{4}$$

The wavelet covariance $P_{\mathbf{w}}$ is rarely a diagonal matrix, indicating that the wavelet coefficients are *not*, in fact, independent. Indeed, it is well known that localized image structures, such as edges, tend to have substantial power across many scales. We have observed [25] that, although the majority of correlations are very close to zero, a relatively significant percentage (10%) of the coefficients are strongly correlated across several scales or across orientations.

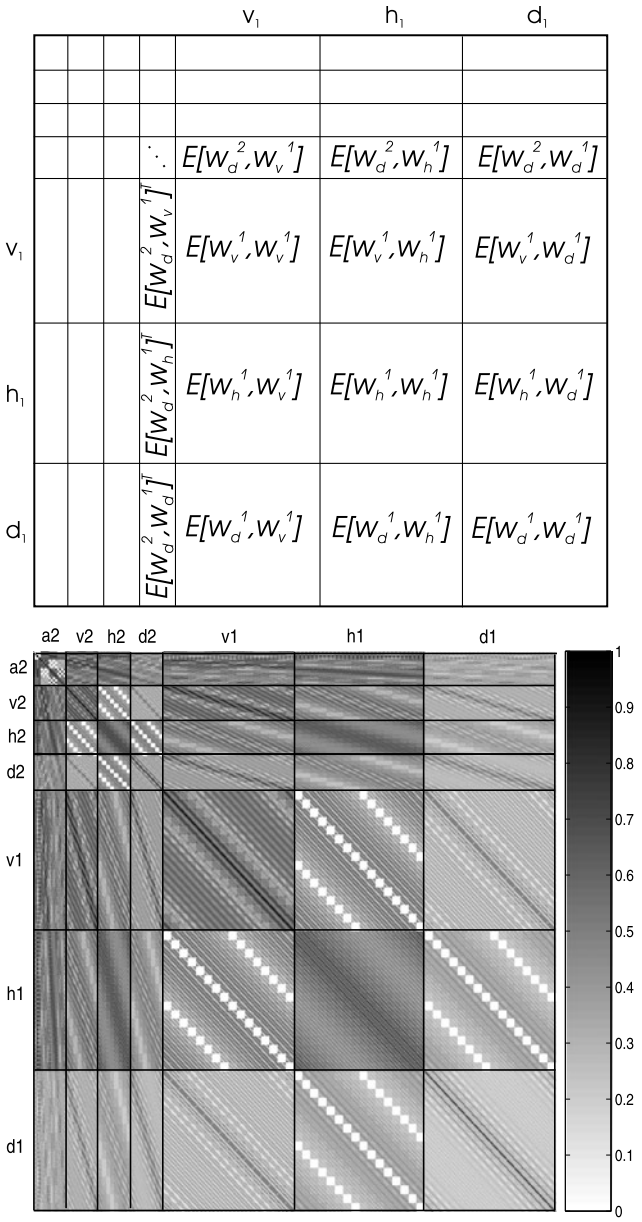


Fig. 4 Wavelet domain covariance matrix $P_{\mathbf{w}}$. *Top*: schematic plot. *Bottom*: Correlation coefficient magnitudes of the spatial thin-plate model (Fig. 3(a)) in the db2 wavelet domain. The main diagonal blocks correspond to the same scale and orientation, whereas off-diagonal blocks illustrate cross-correlations across orientations or across scales

3.1 Wavelet Domain Conditional Histograms

Now we take one step further and consider higher order dependencies. Here, we examine the conditional probability densities of pairs of coefficients, illustrated via conditional histograms (or "butterfly" plots, as in [22]).

For the spatial statistics $P_{\mathbf{x}}$ we consider the thin-plate (Fig. 3(a)), a third-order Gaussian Markov Random Field (GMRF) with the covariance structure shown in Fig. 4. The linearity of the wavelet transform then implies that $\{\mathbf{w}^j\}$ will be jointly Gaussian, as can be seen in Fig. 5, which

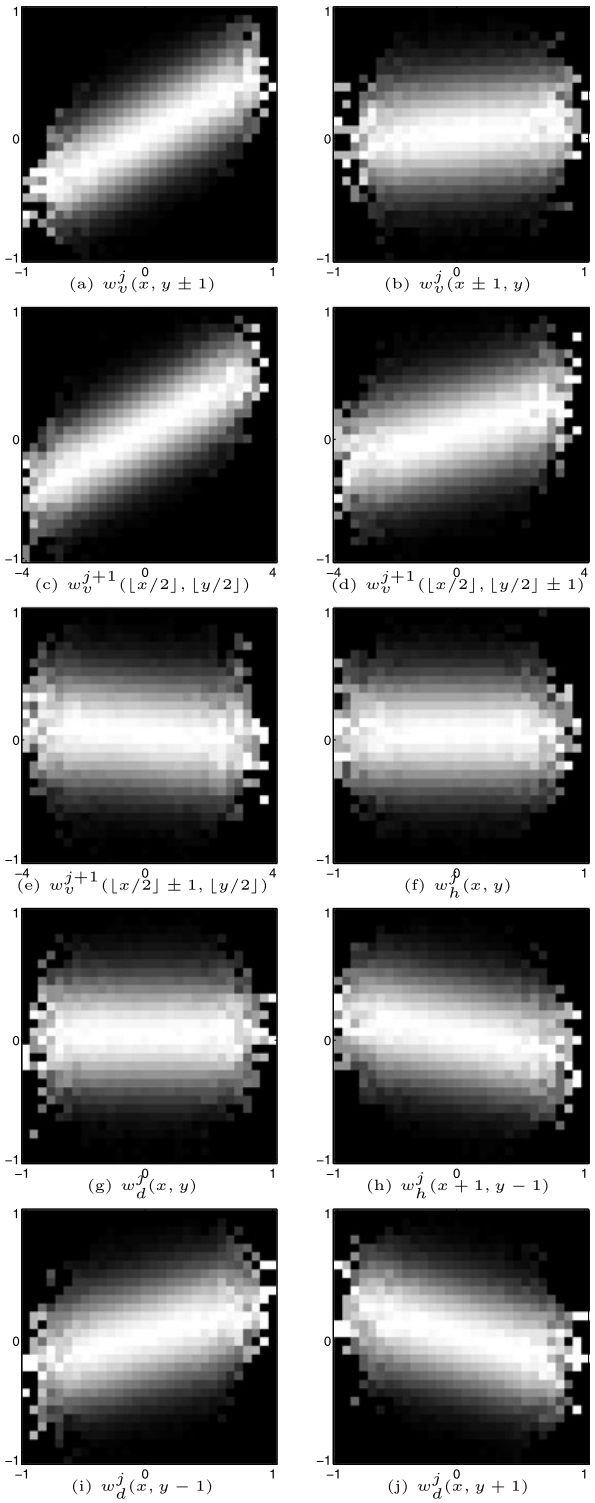


Fig. 5 Conditional histograms of a horizontal coefficient at position $w_v^j(x, y)$ conditioned on coefficients at the same scale and orientation (**a**, **b**), at the same orientation but adjacent scales (**c**–**e**), and at the same scale but across orientations (**f**–**j**). In each plot, brightness indicates probability, with each column being independently rescaled to cover the whole range of intensities. CC shows correlation coefficient value calculated directly from the plot. The spatial domain texture is a GMRF thin-plate decomposed into the db2 basis function

plots the conditional histograms of a single horizontal subband db2 wavelet coefficient $w_v^j(x, y)$ conditioned on eight other coefficients across space, orientation, and scale. These eight coefficients consist of those used in previous works and some new promising ones. These plots highlight the following important aspects:

- Panels (a–b): the correlation direction of two spatially adjacent coefficients is a function of subband: within its subband, a vertical coefficient is positively and more strongly dependent on its vertical neighbors than with its horizontal ones. This observation is intuitive: the row elements in the vertical channel result from the application of a high-pass filter, and are thus less correlated and consequently less dependent than the column elements which result from low-pass filtering.
- Panels (c–e): there is a relationship beyond just the common child-parent assumption, i.e. a child coefficient strongly depends not only on its parent (a fact observed by many researchers) but also on its parent's neighbors.
- Panels (f–g): coefficients at the same location but across different orientations are uncorrelated (although possibly still dependent).
- Panels (h–j): there are inter-orientation correlations, but with pixels at other locations, dependent on the orientation of the associated subband.

Thus, we see confirmed the expected child-parent relationship, but also a strong subband dependence in the spatial correlations.

The assumption of a GMRF prior allowed the simple definition of spatial statistics, however this leads to unrealistically simple wavelet-domain statistics. In order to show that the above patterns are not a direct impact of the GMRF assumption, the wavelet joint statistics were computed empirically over a large class of real images.[1] The panels in Fig. 6 are parallel to those of Fig. 5. Comparing these two sets of histograms one can enumerate these important contradistinctions:

- First and foremost is that the real-image histograms are not linear, while synthetic histograms are. This observation is quite intuitive. Our synthesized texture is the product of a GMRF, an MRF realization whose variable interactions can be represented comprehensively by linear terms. This texture then undergoes a linear wavelet transform that won't introduce non-linear dependencies. More formally, the joint and conditional distributions of GMRF variables and their wavelet transformed peers can be derived to be Gaussian. On the other hand, real image intensities are proven to follow complex non-linear joint distributions; hence their non-linear histograms.

[1] California Institute of Technology CVI Database (http://www.vision.caltech.edu/html-files/archive.html).

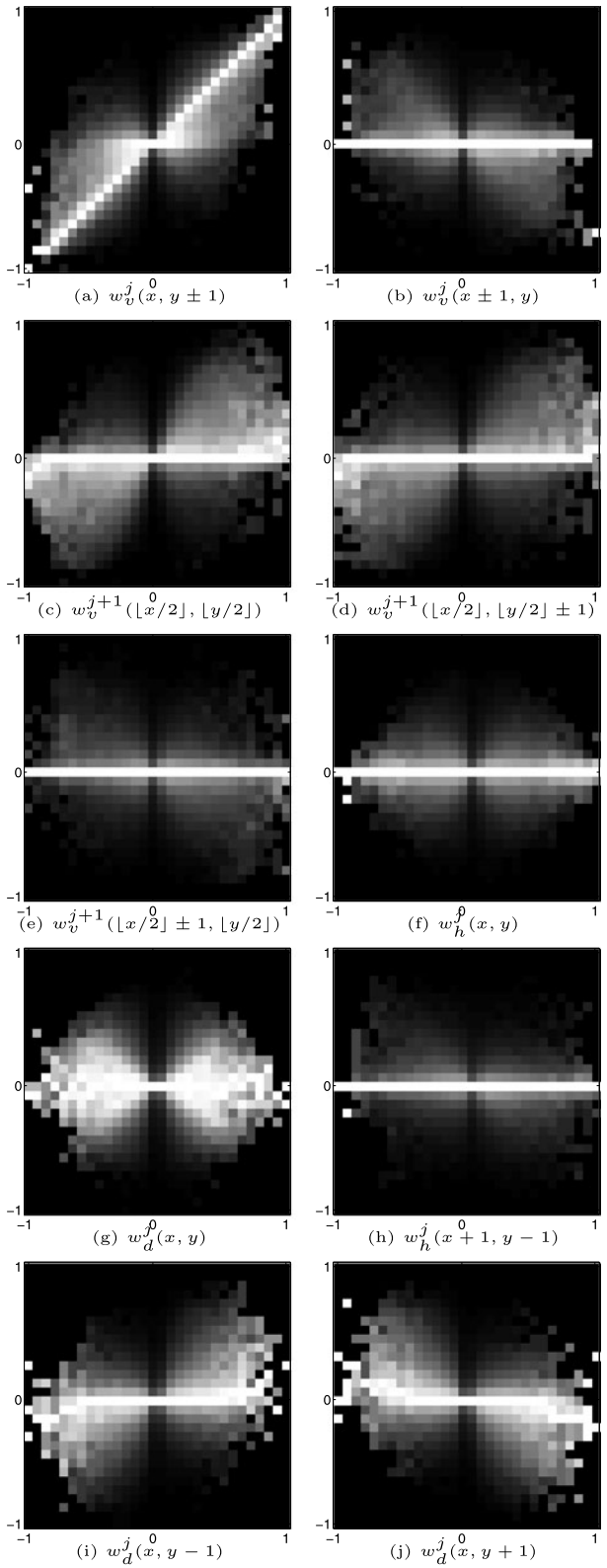
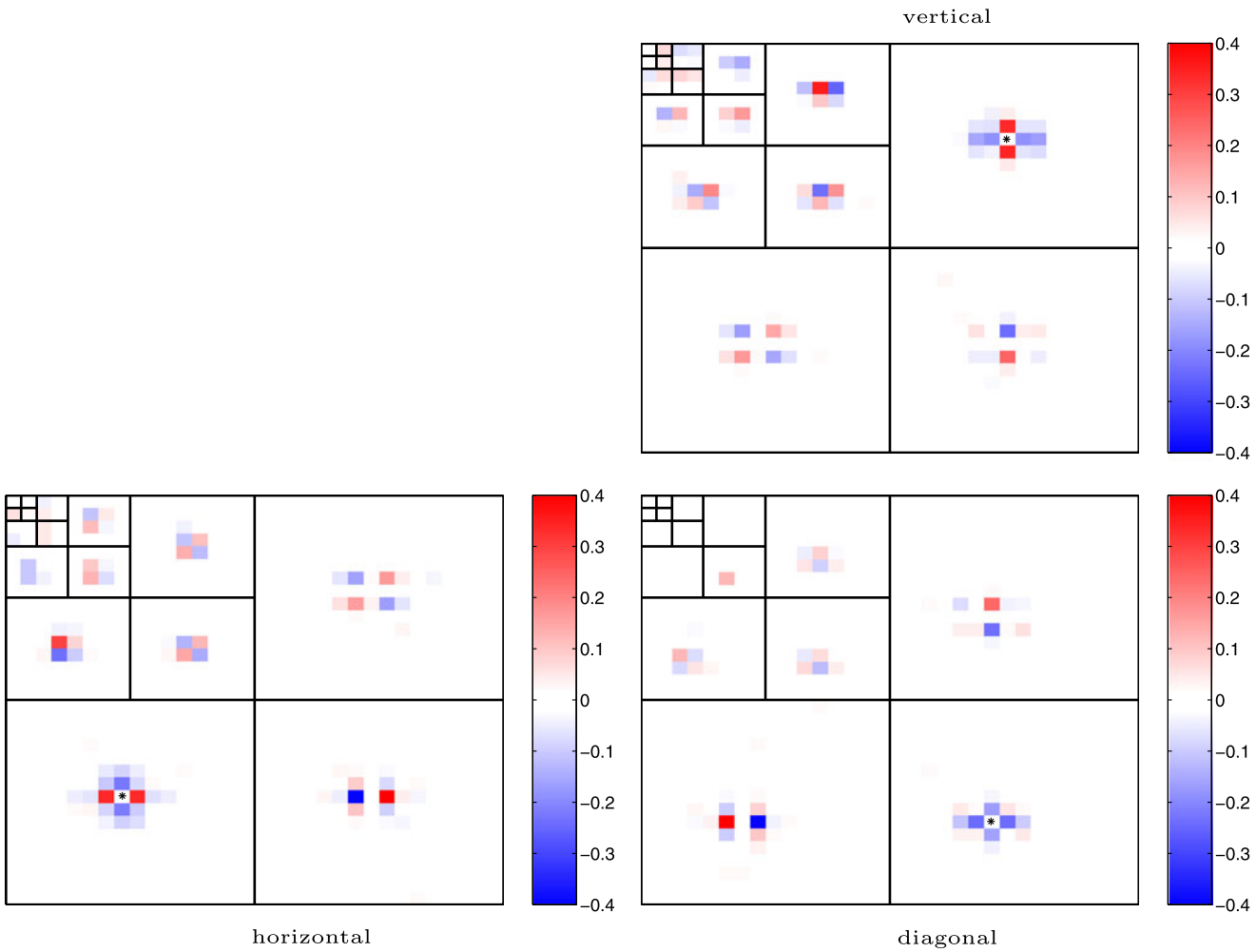


Fig. 6 Conditional histograms of a horizontal coefficient corresponding to the plots in Fig. 5. In each plot, brightness indicates probability, with each column being independently rescaled to cover the whole range of intensities. The associated spatial domain images were a collection of photographs, taken from an image database

– The next inconsistency corresponds to the value zero on the horizontal axis. Real-image histograms have a special sensitivity to the abscissa of zero. As evident in all the ten plots in Fig. 6, a nil value for the conditioning variable dooms the conditioned variable to be nil. To explain this phenomenon we have to note that a real image usually consists of large patches of smooth areas and lengths of sparse edges separating them. In the wavelet domain this translates into a majority of zero coefficients and a minority of adjacent significant coefficients. Smooth areas induce zero coefficients in all orientations and edges bring about non-zero values usually (and not always), in all orientations. As a result, if some particular wavelet coefficient is zero there is a great chance that it has been in a smooth area and so the neighbors are zero too, otherwise it should correspond to a perfect horizontal, vertical or diagonal edge (an edge which causes significant values only in one orientation) which is too scarce because of the superposition each scale of wavelet decomposition brings in. Meanwhile, our synthesized texture histogram is not troubled around zero because none of above arguments hold true for it. As opposed to real images, the GMRF texture is fraught with edges to an extent that smooth areas are restricted to boundaries between positive and negative edges. Consequently, there are much fewer zero coefficients compared to the real case and whenever they occur they are surrounded by significant values. That is why a zero conditioning variable in the GMRF case doesn't limit the conditioned variable. As for a mathematical explanation, we can say, because every GMRF coefficient can be expressed by a linear combination of its neighbors, if one of the contributing neighbors happen to be zero, others may still make up for it. On the other side however, a zero neighbor of a real-image coefficient has probably a multiplicative effect on the coefficient making it restricted to zero.
– Another prominent difference is in the crests. Synthetic histograms have wide blunt crests implying low kurtosis (quite reasonable because of the Gaussian nature) while real-image histograms have very sharp crests suggesting high kurtosis (an observation compliant with the belief established in the literature about wavelet coefficient statistics).

Although the details in Fig. 6 are considerably richer and more complicated, the signs and degrees of correlation are very similar to those in Fig. 5, leading to the same conclusions of subband-dependent structure.

With the conditional distributions of Figs. 5–6 by way of introduction, we see that existing wavelet joint models consider only a *subset* of these inter-relationships; furthermore the two-mixture Gaussian marginals of [1], which have an underlying binary state (high/low variance), will be unable to capture the correlations such as in the distribution of Fig. 6(a).



(a) Correlation Coefficients

Fig. 7 Wavelet (db2) correlation structures averaged over the textures displayed in Fig. 3. Each panel plots the correlation of a selected coefficient (•) with all other coefficients at all orientations and scales (the coefficient's parent is marked by □). The *top panels* show correlation coefficients; the *bottom panels* plot the MSE estimation significance [25] of the interrelationships

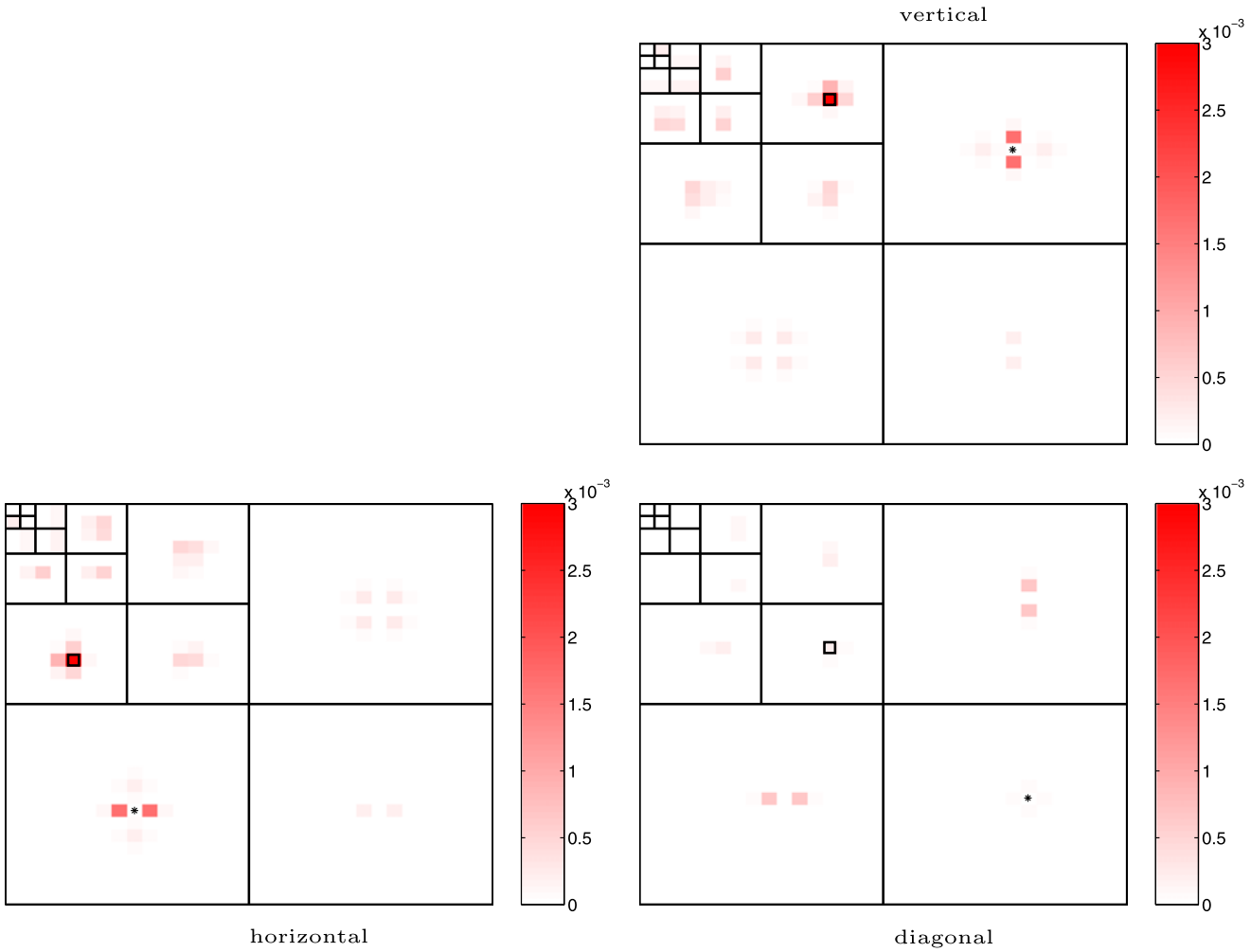
3.2 Wavelet Domain Correlation Structure

Examining wavelet individual joint distributions, such as in Figs. 5–6, precludes a high-level, comprehensive study. Henceforth, we will focus on correlation rather than dependency since the treatment of an elaborate map of non-linear dependency is beyond the scope of this article and needs a thorough study in a separate work.

Although we readily acknowledge the limitations of using correlation as a measure of joint behavior (e.g. the uncorrelated, but dependent behavior of Fig. 6(g)), in general the coefficients correlations in Fig. 6 give a reasonable sense of the joint behavior. That is, the dependence between coefficients are mostly due to correlation. The great advantage of correlation, is the lower complexity the algorithms using it have, letting us opt for a larger (more inclusive) and in turn a more efficient map, the primary contribution of this paper.

Figure 7(a) shows the correlation coefficients, averaged over the four wavelet priors P_w corresponding to the textures of Fig. 3: the observed behaviors are thus persistent patterns, not the peculiar behavior of a single, particular texture. For simplicity of interpretation, we use the traditional 2-D wavelet plot to display the correlations of a selected coefficient with *all* other coefficients on the entire wavelet tree.

Because correlation coefficients can be misleading (a high correlation between two tiny-variance wavelet elements may not be of modeling significance), we also plot the correlation *significance* [25], measured as the reduction in mean squared estimation error induced by including the correlation relationship (Fig. 7(b)). Thus our de-



(b) Correlation Significances

Fig. 7 (*Continued*)

pendence monitoring techniques illustrated important interacting neighbors and took care of necessity of presence of a coefficient. With correlation significance we can show which neighbors will not contribute to a better performance and tell where we should decide our map is sufficient.

There is a very clear consistency between both of these maps and the conclusions reached from the conditional histograms. In particular, the correlation structure is spatially-localized and sparse, and there is clear persistence [11] across scales, basic attributes which nearly all wavelet models have in common. The local neighborhood for any given pixel is not limited to the pixel's subband: every coefficient exhibits correlations extending across multiple scales and strong correlations with spatially near neighbors both within subband and across orientations. The correlation structure for horizontally and vertically aligned coefficients are almost symmetrically identical. There are, however, striking patterns which are *not* reflected in other models:

1. A given coefficient is not linearly correlated with siblings at other orientations, although a statistical dependence may remain, as in Fig. 6(g). Within any particular scale, across-subband siblings are nearly decorrelated, whereas across-subband neighbors of siblings are linearly related.
2. Within-subband correlations are orientation-dependent: vertical-band coefficients are positively-correlated vertically, negatively horizontally, and vice versa for horizontal coefficients.
3. Inter-subband correlations are orientation-dependent: horizontal-band coefficients are correlated with horizontal neighbors in the diagonal subband, vertical coefficients are correlated with vertical neighbors.
4. Inter-scale correlations are orientation-dependent: in addition to its parent, a coefficient is correlated with the

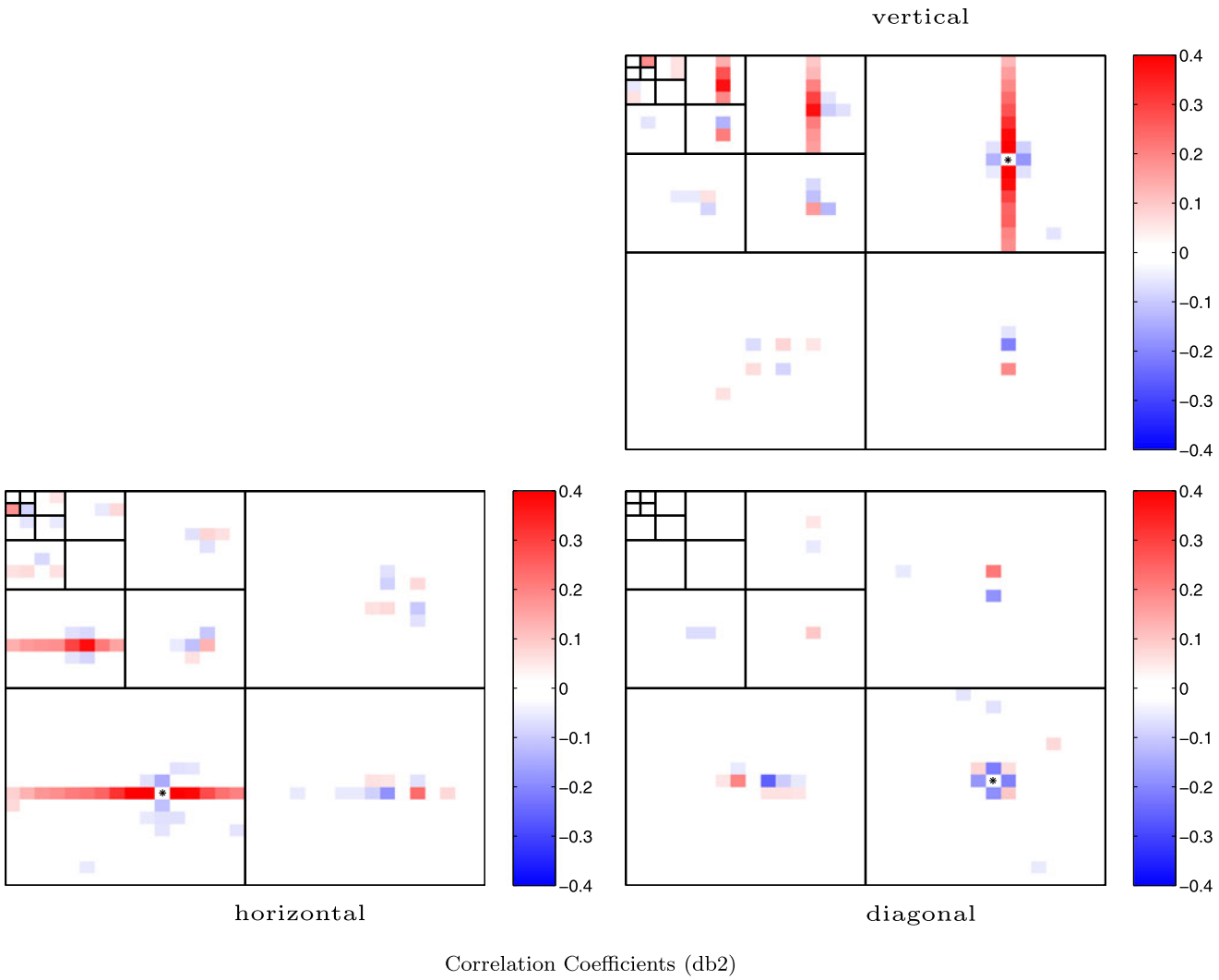


Fig. 8 Wavelet (db2,4) correlation structures averaged over a collection of 5000 real images. Each panel contains three plots illustrating the correlations of a given coefficient (marked by •) with its local neighborhoods in the horizontal, vertical, and diagonal subbands. The standard location of parent is marked by □. Clearly, the location of parents and siblings is a function of the mother wavelet length

spatial neighbors of its parent, e.g. a vertical-band coefficient is positively correlated with vertical neighbors of its parent, and weakly negatively with horizontal neighbors.

Finally, to confirm that our conclusions are not the result of Markovianity, Gaussianity, or other coincidences associated with our choices of textures, Fig. 8 plots the correlation maps for db2 and db4 wavelets averaged over a collection of 5000 randomly cropped and subsampled real-world images. The consistency between Figs. 7–8 is very clear and both panels of Fig. 8 support the essential conclusions, above, of sibling uncorrelatedness and within-subband/inter-subband/inter-scale orientation dependent correlatedness.

Regarding our observations and the trade-off between complexity and accuracy we propose neighborhood structures, already reported in our previous short paper [27], that we shall use in the next section to test for viability. For the sake of completeness, we include the neighborhood structure here too.

For a coefficient w_i belonging to the wavelet coefficients set $\underline{w} = \{\underline{w}_h, \underline{w}_v, \underline{w}_d\}$ we define

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

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

$$s_{ud}(i): \begin{array}{c}\bullet\\ \times\\ \bullet\end{array} \qquad s_{lr}(i): \bullet\times\bullet \qquad s_2(i): \begin{array}{c}\bullet\\ \times\\ \bullet\end{array}$$

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 $s_n(i)$ defines various sibling sets (on the same scale as w_i). This allows us to propose two asymmetric

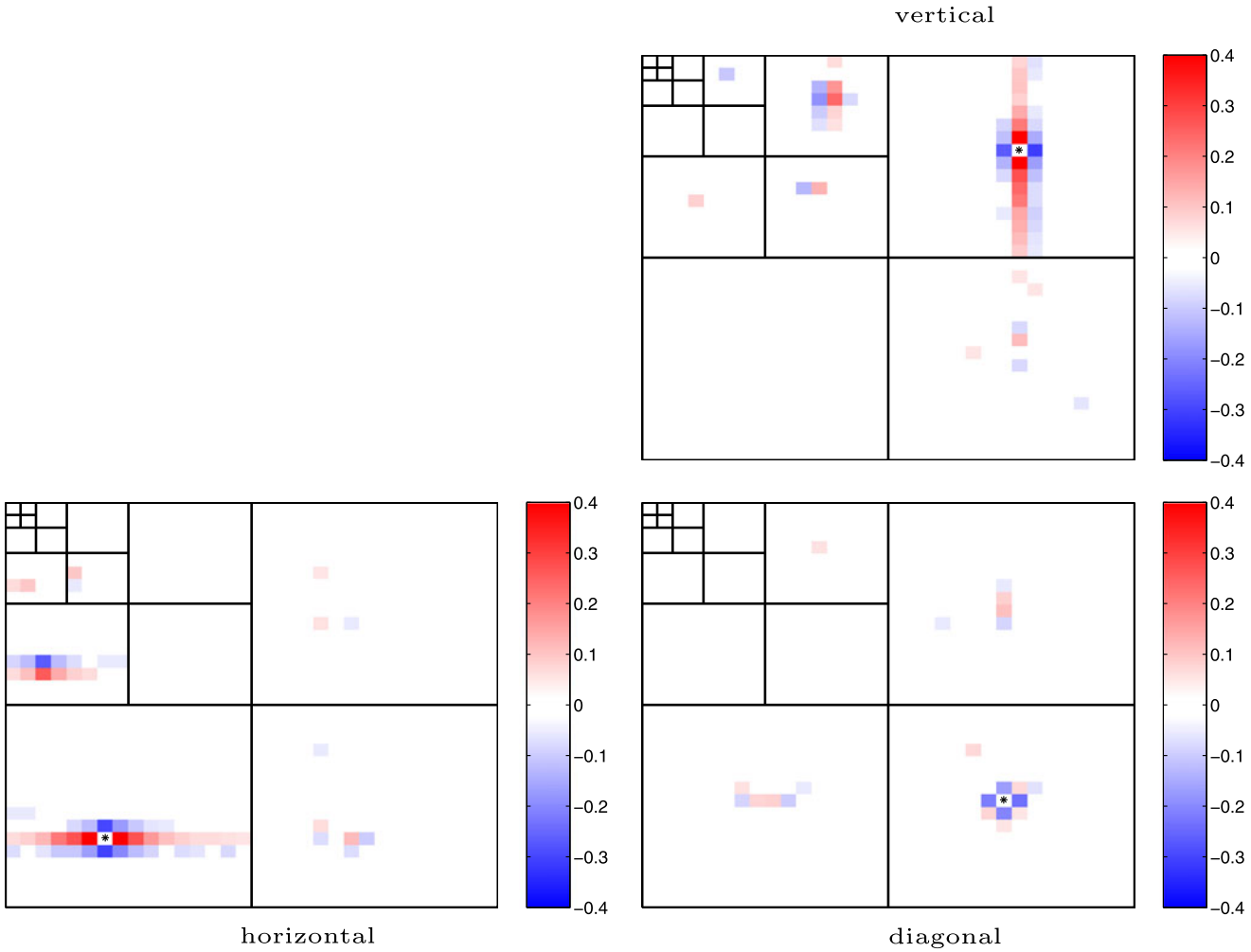


Fig. 8 (*Continued*)

neighborhood structures:

$$\mathcal{N}_1(i) = \begin{cases} \{s_{lr}(i), p_1(i)\}; & w_i \in \underline{w}_h \\ \{s_{ud}(i), p_1(i)\}; & w_i \in \underline{w}_v \\ \{s_2(i), p_1(i)\}; & w_i \in \underline{w}_d \end{cases}$$

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

where operators d, v, and h return diagonal, vertical, and horizontal subband counterparts.

4 Results

In previous sections we studied the correlation structure governing wavelet coefficients of real images and random textures. First we examined sufficiency of including a map of correlation with conditional histograms and correlation coefficients of a sample wavelet coefficient. Then necessity of inclusion of a given correlation between two coefficients was investigated by correlation significance. Finally a neighborhood structure or map was proposed meant to include crucial linear dependence information.

With this correlation map in place, ideal is to adopt it in a probabilistic framework capable of efficiently circulating information through the network and calculating proper estimated coefficient values. However, as pointed out earlier, this should be addressed in a separate paper for the difficulty of handling networks with cycles in their adjacency graphs. Our intention here has been to empirically inspect coefficient dependence and correlation in particular and to introduce a reliable neighborhood which can be used as a basis for future analysis of wavelet coefficients. That being

said, we content ourselves with a result section overhauling the map in two simple yet efficient estimation frameworks.

4.1 Local Linear Estimator

Based on the correlation map of Fig. 8, one can define various different neighborhoods. A typical neighborhood $\mathcal{N}_2$ as defined above is used for the experiments below. Since we put more focus on linear dependence of coefficients throughout the paper, we try out our map in a typical linear estimation framework, Local-Estimate, which is introduced in our previous work [28] and serves only as a means of examining the efficiency of our map:

1. The given random field $\underline{x}$ is projected into the wavelet domain with the resulting coefficient vector $\underline{w}$. Assuming additive disturbance ν we get the observation y:

   $$y_i = w_i + \nu_i$$

   Let us form two neighborhood vectors:

   $$\underline{y}_i = [y_i, \{y_j;\, j \in \mathcal{N}_i\}]^T$$

   $$\underline{w}_i = [w_i, \{w_j;\, j \in \mathcal{N}_i\}]^T$$

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

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

   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.

The performance of the Local-Estimate with the proposed correlation map is compared with that of HMT in several denoising tasks. With this comparison we seek to vindicate the virtue of our map by showing that even with a linear estimator the results are better than a fairly-good non-linear estimator. The behavior of BayesShrink is also set forth as a reputable non-linear independent shrinkage method. By this, we want to recognize the position of our neighborhood system adopted in the Local-Estimate with another famous algorithm in the literature.

In this experiment Goldhill image is corrupted with additive white Gaussian noise with different standard deviations. The performance measure used, is the Root Mean Squared Error (RMSE). The results are plotted in Fig. 9. Obviously the linear Local-Estimate does much better in denoising

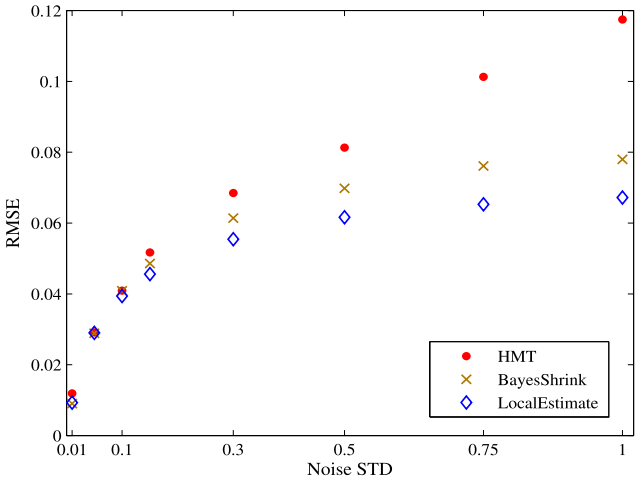


Fig. 9 Denoising performance of the HMT algorithm, BayesShrink and Local-Estimate (using the proposed correlation map) in terms of Root Mean Squared Error

the image as compared to HMT specifically when noise standard deviation is high. The linear method also beats BayesShrink in this experiment. The resulting denoised images from the HMT and Local-Estimate are illustrated in Fig. 10. HMT fails to produce genuine edges and introduces aliasing effects. This is due to the lack of detail coefficients to trim edges. Because HMT decides the significance of a coefficient based on its coarser parent it may eliminate many significant finer-scale values because of a fault coarser-scale coefficient. Also, since there are four children to each coarse coefficient in the quad-tree representation of the HMT is prone to mistakenly encourage insignificant children of a truly significant parent or frustrate significant children of a truly insignificant parent. Although these faults may occur in any dependency-based framework, the problem here is that these faults are always biased to liken children of a parent and hence produce coarser edges. The Local-Estimate image, albeit unbecoming to the eyes, very well follows prominent edges.

Although non-linear denoisers have an essential advantage over the experimented Local-Estimate (as a linear estimator), the latter has the upper hand in time matters. The point here however, is just to verify the fitness of our proposed map and not to introduce a novel denoising algorithm and compare it to the latest algorithms in the field.

It is important to note that, increasing the span of the neighborhood leads to better estimation achievements. But the time and computation costs will grow exponentially only to get dwindling performance gains. According to this philosophy was our map organized, to stay small while employing the most of correlation information.

(a)

(b)

(c)

Fig. 10 Visual comparison of the output of the denoisers on the Goldhill image with $\sigma = 0.15$ noise (**a**). HMT image is (**b**) and Local-Estimate (**c**)

4.2 Correlated Non-linear Shrinkage

We now extend our previous experiment by developing non-linear correlated wavelet shrinkage [28]. As in previous subsection we derive the Local-Estimate output and then use it as a basis for non-linear estimation:

1. The noisy wavelet-domain random field is observed as: $y_i = w_i + v_i$.
2. Wavelet coefficients $\underline{w_i}$ are assumed jointly Gaussian, and we derive the Local-Estimate output.
3. The final estimate $\hat{w}_i$ is found via some non-linear shrinkage method. For the case of state-of-the-art ProbShrink [29],

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

A schematic display of our Correlated Shrinkage algorithm is:

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

As before, 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. 11. 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. 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 (Figs. 11 and 12). There are evident blurring artifacts in the ProbShrink image. This is because ProbShrink is trying to choose the Generalized Laplacian distribution shape parameters in such a way that no noisy coefficient qualifies as significant. But this also eliminates significant (uncorrupt) coefficients since it is judging solely on the value corresponding to each coefficient. That is, since we have a large amount of noise, the threshold for a given coefficient to be original is very high and this bars even those uncorrupt coefficients that are not necessarily prominent but account for the small details in edges. Furthermore, ProbShrink fails to produce uniform edges because it does not consider neighboring coefficients to fine-tune the estimate for a given position while we know different coefficients are differently affected by noise and cannot be treated independently. Therefore, the ProbShrink (or any other independent approach) image both lacks in edge details and nearly bereft of smooth edges. As a result, the mother wavelet (db2) reveals itself in the image because of some sharp changes in the estimated wavelet

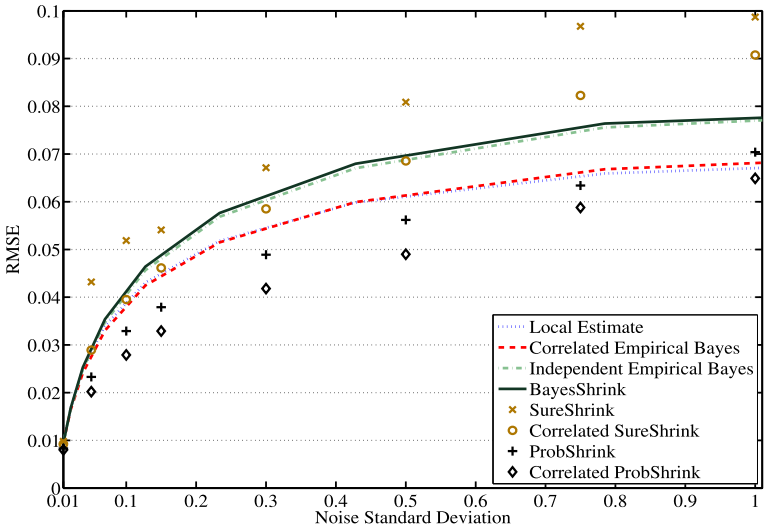


Fig. 11 Comparison of performance of different algorithms in different noise σ in terms of RMSE. As expected, the combined Empirical Bayes, SUREShrink and ProbShrink have improved performances over the corresponding original methods

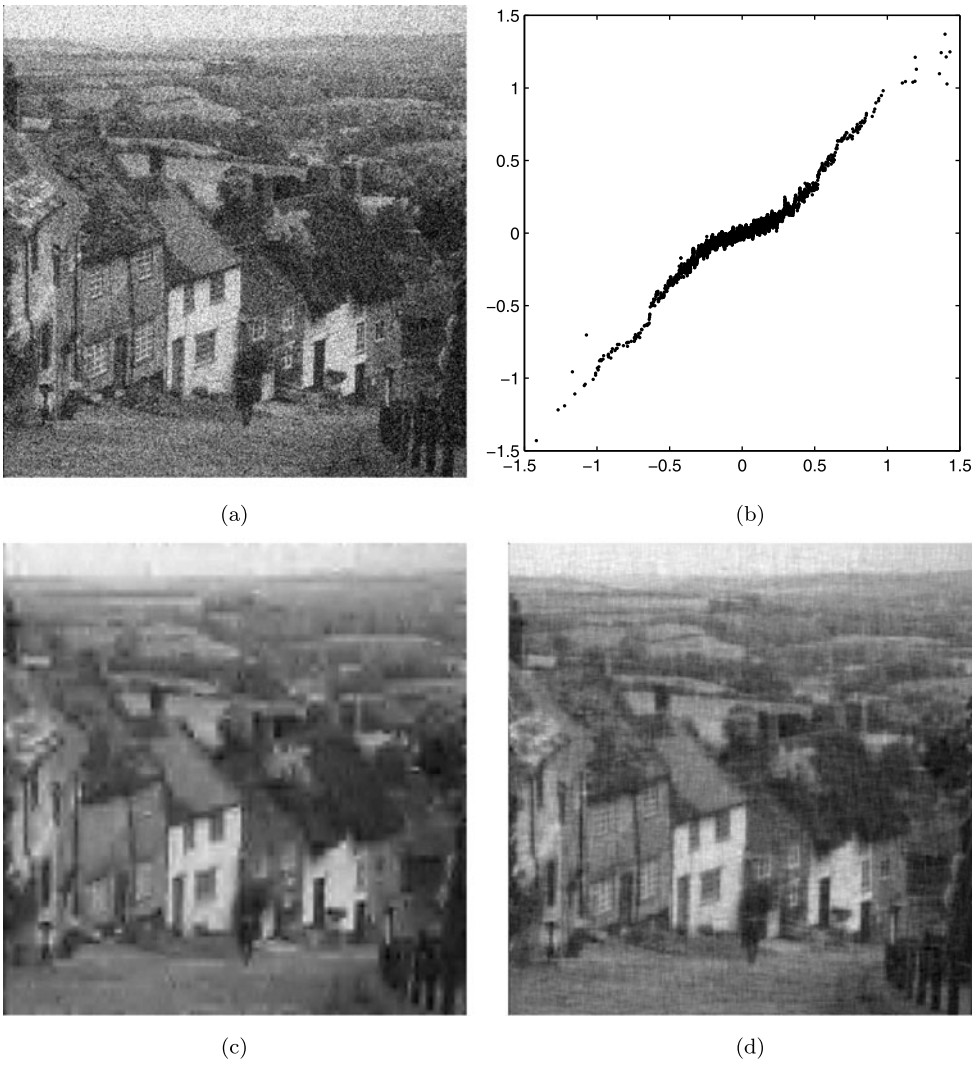


Fig. 12 Experimental results of correlated wavelet shrinkage: (**a**) Noisy image ($\sigma = 0.15$), (**b**) Non-linear transform function of the hybrid algorithm, (**c**) Linear Local Estimate denoised image, (**d**) Combined Local Estimate and ProbShrink denoised image

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. Figure 12(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.

With these observations in place, the advantage of shrinkage algorithms taking advantage of wavelet correlations both the virtue of our neighborhood system have been verified. In particular, there are some striking improvements in Fig. 11 which merit further study. An interesting research direction is to incorporate the correlation structures of Figs. 7–8 into promising HMT's [2, 10, 11]. 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.

5 Conclusions

A thorough study of empirical 2-D wavelet correlations has been presented in this paper. The expected patterns—correlation sparsity and parent-child persistence—are clear, however there are additional striking relationships which, as yet, are not normally found in wavelet models. In particular, coefficients are very nearly decorrelated with siblings across orientations, however, there is a very strong orientation-dependence governing correlations within subbands, across orientations, and across scales. To address the primary motivation of this study, the proposed correlation structures can be considered as a map to refine other wavelet-domain probabilistic models, such as HMMs, and to examine those model's neighborhood sufficiency and necessity.

Since most of the state-of-the-art denoising algorithms consider dependencies between wavelet domain coefficients, a careful study of dependencies (instead of correlations) in a neighborhood of a coefficient seems indispensable. In a future work we will address this issue to find a map of dependencies in a wavelet tree.

The inefficiency of the 2D wavelet transform in detecting one-dimensional image singularities (edges) , in addition to point singularities, has led to a new generation of multiresolution representations including Contourlet [16] and Shearlet [30] transforms, which combine ideas of multiscale analysis and directional filtering in transform design. These frameworks result in many subbands (more than the three subbands associated with the classical wavelet transform), adapting in scale, position and orientation to the edges. Because of this multitude of subbands and since these transforms are overcomplete, there exists a significant amount of across-subband correlations. Statistics of these recent transformations can be examined by extensions of our wavelet modeling taking into account their individual idiosyncrasies. Although, there are already a couple of works addressing coefficient joint statistics analysis from different points of view, such as those considering Curvelets [13–15] or Contourlets [16], our work would be different in terms of the goals and means. We will try to find relation maps usable in different correlation-based and graph-based estimation frameworks using a conclusively large real image set and cross examining each neighborhood position by its significance measure.

References

1. Chipman, H., Kolaczyk, E., McCulloch, R.: Adaptive Bayesian wavelet shrinkage. J. Am. Stat. Assoc. 92–99 (1997)
2. Crouse, M.S., Nowak, R.D., Baraniuk, R.G.: Wavelet-based statistical signal processing using hidden Markov models. IEEE Trans. Signal Process. **46**, 886–902 (1998)
3. Liu, J., Moulin, P.: Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients. IEEE Trans. Image Process. **10**, 1647–1658 (2001)
4. Abramovich, F., Sapatinas, T., Silverman, B.W.: Wavelet thresholding via a Bayesian approach. J. R. Stat. Soc. B **60**, 725–749 (1998)
5. Moulin, P., Liu, J.: Analysis of multiresolution image denoising schemes using a generalized Gaussian and complexity priors. IEEE Trans. Inf. Theory **45**, 909–919 (1999)
6. Shapiro, J.: Embedded image coding using zerotrees of wavelet coefficients. IEEE Trans. Signal Process. **41**, 3445–3462 (1993)
7. Malfait, M., Roose, D.: Wavelet-based image denoising using a Markov random field a priori model. IEEE Trans. Image Process. **6**, 549–565 (1997)
8. Pizurica, A., Philips, W., Lemahieu, I., Acheroy, M.: A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising. IEEE Trans. Image Process. **11**, 545–557 (2002)
9. Srivastava, A.: Stochastic models for capturing image variability. IEEE Signal Process. Mag. **19**(5), 63–76 (2002)
10. Fan, G., Xia, X.: Wavelet-based texture analysis and synthesis using hidden Markov models. IEEE Trans. Circuits Syst. **50**, 106–120 (2003)
11. Romberg, J., Choi, H., Baraniuk, R.: Bayesian tree-structured imaged modeling using wavelet-domain hidden Markov models. IEEE Trans. Image Process. **10**, 1056–1068 (2001)
12. Portilla, J., Strela, V., Wainwright, M.J., Simoncelli, E.: Image denoising using Gaussian scale mixtures in the wavelet domain. IEEE Trans. Image Process. **12**, 1338–1351 (2003)
13. Boubchir, L., Fadili, J.M.: Multivariate statistical modeling of images with the curvelet transform. In: Proc. 8th Inter. Symposium on Signal Processing and Its Applications (ISSPA 2005), pp. 747–750 (2005)
14. Alecu, A., Munteanu, A., Pizurica, A., Philips, W., Cornelis, J., Schelkens, P.: Information-theoretic analysis of dependencies between curvelet coefficients. In: ICIP, pp. 1617–1620 (2006)
15. Tessens, L., Pizurica, A., Alecu, A., Munteanu, A., Philips, W.: Context adaptive image denoising through modeling of curvelet domain statistics. J. Electron. Imag. **17**(3) (2008). doi:10.1117/1.2987723

16. Po, D.D.-Y., Do, M.N.: Directional multiscale modeling of images using the contourlet transform. IEEE Trans. Image Process. **15**(6), 1610–1620 (2006)
17. Liu, J., Moulin, P.: Information–theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients. IEEE Trans. Image Process. 1647–1658 (2001)
18. Xu, Y., Weaver, J., Healy, D., Lu, J.: Wavelet transform domain filters: a spatially selective noise filtration technique. IEEE Trans. Image Process. **3**(6), 747–758 (1994)
19. Fan, G., Xia, X.: Improved hidden Markov models in the wavelet-domain. IEEE Trans. Signal Process. **49**, 115–120 (2001)
20. Mihcak, M., Kozintsev, I., Ramchandran, K.: Low-complexity image denoising based on statistical modeling of wavelet coefficients. IEEE Signal Process. Lett. **6**, 300–303 (1999)
21. Nowak, R.D.: Multiscale hidden Markov models for Bayesian image analysis. In: Vidakovic, B., Muller, P. (eds.) Wavelet Based Models
22. Simoncelli, E.: Modeling the joint statistics of images in the wavelet domain. In: Proceedings of the SPIE 44th Annual Meeting, pp. 188–195 (1999)
23. Chang, S., Yu, B., Vetterli, M.: Spatially adaptive wavelet thresholding with context modeling for image denoising. IEEE Trans. Image Process. **9**, 1522–1531 (2000)
24. Vannucci, M., Corradi, F.: Covariance structure of wavelet coefficients: theory and models in a Bayesian perspective. J. R. Stat. Soc. B **61**, 971–986 (1999)
25. Azimifar, Z., Fieguth, P., Jernigan, E.: Towards random field modeling of wavelet statistics. In: Proceedings of the 9th ICIP (2002)
26. Azimifar, Z., Fieguth, P., Jernigan, E.: Hierarchical Markov models for wavelet-domain statistics. In: Proceedings of the 12th IEEE Statistical Signal Processing Workshop (2003)
27. Azimifar, Z., Fieguth, P., Jernigan, E.: Correlated wavelet shrinkage: models of local random fields across multiple resolutions. In: Proceedings of the 12th ICIP (2005)
28. Amiri, M., Azimifar, Z., Fieguth, P., Jernigan, E.: Correlated nonlinear wavelet shrinkage. In: Proceedings of the 15th ICIP (2008)
29. Pizurica, A., Philips, W.: Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising. IEEE Trans. Image Process. **15**(3), 654–665 (2006)
30. Easley, G., Labate, D., Lim, W.-Q.: Sparse directional image representations using the discrete shearlet transform. Appl. Comput. Harmon. Anal. **25**(1), 25–46 (2008)

Zohreh Azimifar received B.Sc. degree in computer science and engineering from Shiraz University, Shiraz, Iran, in 1994 and her Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, Canada, in 2005. In 2005, she was a postdoctoral fellow in medical biophysics at the University of Toronto, Toronto, Canada. Since 2006 she has been a faculty member and director of computer vision and pattern recognition lab in computer science and engineering at Shiraz University. Her research interest includes statistical learning of generative and discriminative models for massive data sources with particular attention in statistical computer vision and pattern recognition.

Mehdi Amiri graduated from the department of Computer Science and Engineering, Shiraz University, in 2010 with a B.Sc. in Computer Hardware and M.Sc. in Artificial Intelligence. He is a researcher with the Vision Laboratory of Shiraz University and favors fields like Machine Vision, Pattern Recognition and Data Compression.

Paul Fieguth received the B.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, in 1991 and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1995, both in electrical engineering. He joined the faculty at the University of Waterloo in 1996, where he is currently Professor and chair in Systems Design Engineering. He has held visiting appointments at the Cambridge Research Laboratory, at Oxford University, Oxford, UK, and the Rutherford Appleton Laboratory, UK, and at INRIA/Sophia, France,with postdoctoral positions in Computer Science at the University of Toronto and in Information and Decision Systems at MIT. His research interests include statistical signal and image processing, hierarchical algorithms, data fusion, and the interdisciplinary applications of such methods, particularly to remote sensing.

Ed Jernigan received his B.S., M.S., and Ph.D. degrees in Electrical Engineering from the Massachusetts Institute of Technology in 1969, 1971, and 1975, respectively. He joined the Department of Systems Design Engineering at the University of Waterloo in 1976, where he is currently Professor and past Chair. He is a Distinguished Teacher of the University of Waterloo. His research interests include non-linear and adaptive systems for signal and image processing, vision and machine perception, and pattern recognition, particularly with applications in medical imaging and remote sensing. Since 1984 he has served as the Program Director for the Shad Valley summer enrichment program. In 2004 he launched a new university wide enrichment program for high school students of exceptional potential, Waterloo Unlimited.