PERGAMON

Pattern Recognition 36 (2003) 1737–1746

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Hybrid inter- and intra-wavelet scale image restoration

Lei Zhang[a], Paul Bao[b,*], Xiaolin Wu[c]

[a] *Department of Computing, The Hong Kong Polytechnic University, Hung Hum, Kowloon, Hong Kong*
[b] *Department of Information Engineering, The Chinese University of Hong Kong, Shatin, New Territory, Hong Kong*
[c] *Department of Computer Science, University of Western Ontario, London, Ont., Canada*

Received 15 February 2002; received in revised form 18 November 2002; accepted 18 November 2002

Abstract

This paper exploits both the inter- and intra-scale interdependencies that exist in wavelet coefficients to improve image restoration from noise-corrupted data. Using an over-complete wavelet expansion, we group the wavelet coefficients with the same spatial orientation at several scales. We then apply the linear minimum mean squared-error estimation to smooth noise. This scheme exploits the inter-scale correlation information of wavelet coefficients. To exploit the intra-scale dependencies, we calculate the co-variance matrix of each vector locally using a centered square-shaped window. Experiments show that the proposed hybrid scheme significantly outperforms methods exploiting only the intra- or inter-scale dependencies. The performance of noise removal also depends on wavelet filters. In our experiments a biorthogonal wavelet, which best characterizes the image inter-scale dependencies, achieves the best results.

Keywords: Image restoration; Overcomplete wavelet expansion; Inter- and intra-scale dependency; LMMSE

1. Introduction

Wavelet transform (WT) [1–4] can decorrelate images into nearly independent coefficients and is a powerful tool in adaptive signal processing such as coding and noise removal. In the wavelet domain, image energy is compacted into a small number of significant coefficients that represent instantaneous features. The small valued coefficients tend to be localized due to image-independent additive noise, which will increase the image power in general. Wavelet thresholding has proved to be an effective denoising technique [5–11]. In such methods, a threshold is preset to eliminate noise and insignificant structures. Those coefficients less than the threshold will be set to 0 while coefficients above the threshold will be preserved (hard thresholding) or reduced (soft thresholding). After thresholding, the energy of the noisy image may decrease. Generally, the wavelet coefficients within each subband can be modeled as independent identically distributed (i.i.d) random variables with generalized *Gaussian* distribution (GGD) [1,5,6,12]. With the GGD modeling, Chang et al. [5] proposed the *BayesShrink* that employs a subband adaptive soft threshold. Chang's approach significantly outperforms Donoho's classical *Wavelet Shrinkage* [7], which is a global nonparametric estimation method.

Although WT well decorrelates images, there still exist strong dependencies between wavelet coefficients. Exploitation of such dependency information with proper statistical models could further improve the performance of coding and denoising algorithms. As noted by Liu and Moulin [13], the statistical wavelet models can be classified into intra-scale models, inter-scale models and composite dependency models.

A number of intra-scale models have been proposed in Refs. [6,14,15]. Chang et al. [6] defined each wavelet coefficient as a mixture of GGD variables with unknown slowly spatially varying parameters. The estimation of these parameters is conditioned on a function of its neighboring coefficients. The morphological coder [15] also exploits the

* Corresponding author. Tel.: +852-260-98452; fax: +852-260-35032.
E-mail address: paulbao@ie.cuhk.edu.hk (P. Bao).

0031-3203/03/$30.00 © 2003 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0031-3203(02)00350-3

spatial clustering of wavelet coefficients. Mıhçak et al. [14] estimated the two-order local statistics of each coefficient with a centered square-shaped window and developed a linear minimum mean squared-error estimation (LMMSE) like method.

The wavelet dependency information is distributed not only in the intra-scale manner but also in the inter-scale manner. If a coefficient at a coarser scale has small magnitude, its descendant coefficients at finer scales are also likely to be small. Shapiro exploited this property to develop the well-known embedded zerotree wavelet coder [16]. Conversely, if a wavelet coefficient produced by a true signal is of large magnitude at finer scale, its parents at coarser scales will also be large. However for those coefficients produced by noise, the magnitude will decay rapidly along the scales. With this observation, Xu et al. [17] multiplied the adjacent wavelet scales to sharpen the important structures while reducing noise. They then distinguished edge points from noise in the multiscale products. The inter-scale dependencies in the wavelet domain have also been represented by autoregressive models and hidden Markov models in literature [18–21].

Wavelet intra- and inter-scale dependency information can be combined to improve noisy coefficients estimation results. In Ref. [22], Portilla et al. modeled each coefficient as the product of a *Gaussian* random vector and a hidden multiplier variable that included adjacent scales in the conditioning local neighborhood. Liu and Moulin [23] also employed a composite denoising model. Furthermore they analyzed the dependency that existed in wavelet coefficients with a measurement called mutual information [13]. Various wavelet models were also compared with respect to the capability of capturing the dependency information.

In this paper, we will present an adaptive inter- and intra-scale hybrid noise removal scheme with an overcomplete wavelet expansion (OWE). This expansion has been observed to perform much better in denoising than a (bi)orthogonal wavelet transform (OWT) that is a lack of translation invariant [6,9,24]. No downsampling occurs in OWE. Thus at each scale the image samples are the same. We combined the wavelet coefficients at the same position across adjacent scales as a vector, and then applied the LMMSE to the vector variable. This incorporates naturally the inter-scale dependencies of wavelet coefficients to the estimation. To exploit the intra-scale dependency, we estimated the second order statistics of the vector variable locally by a centered square-shaped window within each subband. An LMMSE algorithm with a two-scale composite dependency model is finally proposed and validated with benchmark images. The performance of the scheme also depends on the selection of the wavelet filters.

The paper is organized as follows. Section 2 introduces the OWT and OWE. In Section 3, a hybrid wavelet inter- and intra-scale model is proposed and an efficient denoising algorithm is developed. The effectiveness of several wavelet filters for the algorithm is also discussed. Experiments on benchmark images by intra-scale, inter-scale and the proposed hybrid schemes are shown in Section 4. The conclusion is given in Section 5.

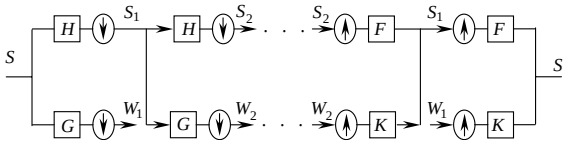


Fig. 1. The decomposition and reconstruction structures of (bi) OWT. The downward and upward arrows indicate downsampling and upsampling.

2. Wavelet transform and overcomplete expansion

A WT represents a signal f as a linear combination of elementary atoms or building blocks. More details about the theory of wavelets and their application in signal processing can be found in Daubechies [3], Mallat [1,2] and Vetterli [4]. Denoted by $\psi_{m,n}$ the dyadic dilation and translation of a mother wavelet ψ can be described as

$$\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n) \tag{2.1}$$

with $m, n \in Z$. Then f can be written as

$$f = \sum c_{m,n}(f)\psi_{m,n}. \tag{2.2}$$

For orthogonal wavelets,

$$c_{m,n}(f) = \langle f, \psi_{m,n}\rangle = \int f(t)\psi_{m,n}(t)\,\mathrm{d}t, \tag{2.3}$$

where $\langle\cdot,\cdot\rangle$ is the inner product in $L_2(\Re)$. For bi-orthogonal wavelets,

$$c_{m,n}(f) = \langle f, \tilde{\psi}_{m,n}\rangle, \tag{2.4}$$

where $\tilde{\psi}_{m,n}$ is the dual wavelet of $\psi_{m,n}$ [3].

The (bi) OWT structure is shown in Fig. 1. The downward and upward arrows indicate downsampling and upsampling. H and G are the low- and high-pass analytic filters while F and K are the corresponding synthetic filters. When the wavelet bases are orthogonal, F and K are the conjugated filters of H and G. Besides the Haar wavelet, there is no other compactly supported orthogonal wavelet that is (anti-)symmetrical. This is a very important property in signal processing. Compactly supported bi-orthogonal wavelets discard the orthogonality to preserve the symmetric property.

OWT is translation variant due to the downsampling in decomposition. This will cause some visual artifacts (such as Gibbs phenomena) in threshold-based denoising [24]. It has been observed that the overcomplete wavelet expansion (OWE, undecimated WT or translation invariant transform) achieves better results in noise removal [6,9,24]. The denoised image by OWE is an average of several circularly shifted denoised versions of the same image by OWT. The noise is smoothed more and the artifacts are weakened.

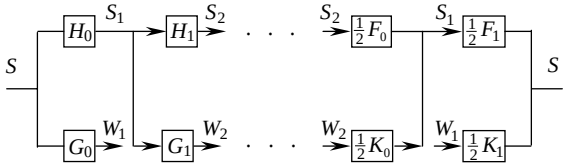


Fig. 2. The illustration of overcomplete wavelet expansion (OWE). Filter H_j is obtained by placing $(2^{j-1}-1)$ zeros between the coefficients of H_0.

The denoising scheme presented in this paper is based on OWE whose structure is shown in Fig. 2. No downsampling or upsampling occurs but the analytic and synthetic filters are changed in each decomposition or reconstruction stage. H_j is obtained by placing $(2^{j-1}-1)$ zeros between the coefficients of H_0. Similarly for G_j, F_j and K_j. The zero padding of wavelet filters, instead of downsampling wavelet coefficients, is used to decrease the frequency bandwidth.

The 2-D OWT and OWE can be extended from 1-D by separable filtering. Wavelet coefficients in the detail bands can be obtained in three directions: horizontal, vertical and diagonal. Fig. 3 shows the one-stage decomposition structures of 2-D OWT and OWE. Filter F' is the transition of F. At each scale the wavelet coefficients of OWE have the same number of samples as the input image.

3. Hybrid intra- and inter-scale model and denoising algorithm

3.1. The LMMSE of wavelet coefficients

Suppose the original image f is corrupted with additive Gaussian white noise ε

$$g = f + \varepsilon, \tag{3.1}$$

where $\varepsilon \in N(0, \sigma^2)$. By implementing OWE to the noisy image, we can denote the wavelet coefficient at scale j as

$$Z_j = X_j + V_j, \tag{3.2}$$

where X_j is the OWE of f and V_j is the OWE of ε for the linearity of the WT.

Many threshold-based methods have been proposed to recover the image from noisy observation. The *Wavelet Shrinkage* $\eta_t(z) = \mathrm{sgn}(z)\max(|z| - t, 0)$ was originally proposed by Donoho [7] with a soft threshold $t = \sigma\sqrt{2\log N}$ based on orthonormal wavelet bases. N is the sample length of input signal. Although Donoho demonstrated it is asymptotically optimal in the minimax sense, it is known that the universal threshold would over-smooth signals. Donoho improved his work by the *SureShrink* scheme in Ref. [8], which adaptively computes the threshold by minimizing Stein's unbiased risk estimate. Bayesian estimation based soft threshold was also proposed in some publications [5,10].

It is generally accepted that the wavelet coefficients conform to the generalized GGD within each subband [1,5,6,12]. With the GGD modeling, Chang et al. [5] presented a subband-dependent soft threshold $t = \sigma^2/\sigma_{X_j}$ (assuming the wavelet base is orthonormal), where σ_{X_j} is the standard deviation of wavelet coefficients X_j. The so-called *BayesShink* scheme outperformed many other popular thresholds.

In this paper, we apply the LMMSE, instead of soft thresholding, to wavelet coefficients. Suppose the variances of V_j and X_j are σ_j^2 and $\sigma_{X_j}^2$, respectively, the LMMSE of variable x_j is

$$\hat{x}_j = cz_j \tag{3.3}$$

with

$$c = \frac{\sigma_{X_j}^2}{\sigma_{X_j}^2 + \sigma_j^2}. \tag{3.4}$$

Noise V_j is Gaussian distributed and independent of X_j. If X_j is also of Gaussian distribution, it is well known that LMMSE will evolve as the optimal minimum mean square-error estimation (MMSE) [25]. Unfortunately X_j subjects to GGD, which degenerates to the Gaussian distribution only in a very special case [1,5].

The standard deviation of noise V_j is computed as

$$\sigma_j = \|\psi_j\|\sigma, \tag{3.5}$$

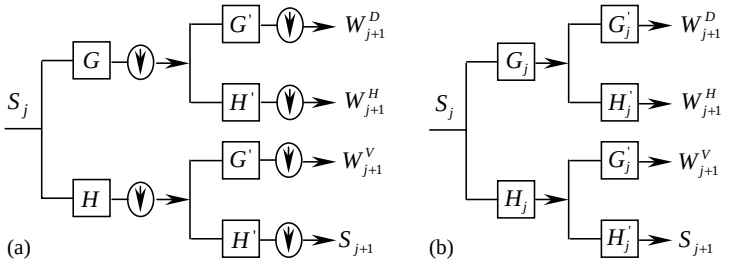


Fig. 3. One stage decomposition of the 2-D (a) OWT; (b) OWE. F' is the transition of F. W_j^H, W_j^V and W_j^D are the wavelet coefficients at horizontal, vertical and diagonal directions.

where ψ_j is the dilation of 2-D mother wavelet ψ: $\psi_j(x,y) = 2^{-j}\psi(2^{-j}x, 2^{-j}y)$, and $\|\bullet\|$ is the normal operator: $\|\psi\| = \sqrt{\iint \psi(x,y)\,\mathrm{d}x\,\mathrm{d}y}$.

If the noiseless image X_j is unknown, σ_{X_j} can be estimated as follows:

$$\hat{\sigma}^2_{X_j} = \sigma^2_{Z_j} - \sigma^2_j \tag{3.6}$$

with

$$\sigma^2_{Z_j} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} Z_j^2(m,n). \tag{3.7}$$

This is the variance of noisy image Z_j whose size is $M \times N$.

LMMSE is similar to the soft thresholding strategy to some extent. Unlike shrinking noisy wavelet coefficient z with a threshold t: $\hat{x} = \mathrm{sgn}(z)\max(|z| - t, 0)$, LMMSE modifies the coefficient with a factor c: $\hat{x} = cz$. It could be observed from Eq. (3.4) that c is less than 1, and then the magnitude of estimated wavelet coefficient $\hat{x}$ will be less than that of z. This would lead to a power reduction of the restored image, similar to the threshold-based schemes.

3.2. The intra-scale dependencies exploited model

The LMMSE scheme presented in Section 3.1 is subband adaptive, but the spatial dependencies within each subband are not exploited. It is usually true that a wavelet coefficient is better correlated with its local neighborhood. Estimating the data locally is widely used technique in image compression and denoising algorithms. In Ref. [6], Chang et al. proposed a spatially adaptive wavelet thresholding scheme based on context modeling. Each wavelet coefficient is modeled as a mixture of GGD variables with unknown slowly spatially varying parameters. The estimation of these parameters is conditioned on a function of its neighboring coefficients.

Mıhçak et al. [14] developed a spatially adaptive scheme with LMMSE strategy. For each wavelet coefficient, the variance $\sigma^2_{X_j}$ is estimated locally with a centered square window Λ_j rather than the whole subband. All the points within Λ_j are assumed to have the same variance $\sigma^2_{\Lambda_j}$, which is calculated as

$$\hat{\sigma}^2_{\Lambda_j} = \max\left(0, \frac{1}{N_{\Lambda_j}}\sum_{(m,n)\in\Lambda_j} Z_j^2(m,n) - \sigma^2_j\right), \tag{3.8}$$

where N_{Λ_j} is number of points in window Λ_j. The modification factor c of the data in window Λ_j is then computed as

$$c = \frac{\hat{\sigma}^2_{\Lambda_j}}{\hat{\sigma}^2_{\Lambda_j} + \sigma^2_j}. \tag{3.9}$$

Mıhçak et al. [14] also imposed an exponential prior distribution $f_\sigma(\sigma^2) = \lambda \mathrm{e}^{-\lambda\sigma^2}$ on the variance $\sigma^2_{\Lambda_j}$. Thus an approximate maximum a posteriori (MAP) estimator of $\sigma^2_{\Lambda_j}$ is

$$\hat{\sigma}^2_{\Lambda_j} = \max\left(0, \frac{N_{\Lambda_j}}{4\lambda}\left[-1 + \sqrt{1 + \frac{8\lambda}{N^2_{\Lambda_j}}\sum_{(m,n)\in\Lambda_j} Z_j^2(m,n)}\right] - \sigma^2_j\right). \tag{3.10}$$

The estimation described in Eq. (3.10) was achieved using a more satisfying denoising performance than that of Eq. (3.8). When the size of window Λ_j increases, the two estimations are identical asymptotically. Mıhçak's scheme exploits the wavelet intra-scale dependencies and its noise removal performance was superior to that of Chang [6].

3.3. The inter-scale dependencies exploited model

Wavelet coefficient dependencies exist not only within but also across subbands. The adjacent wavelet scales are strongly correlated in general. Observing that small magnitude coefficients at coarser scales are more likely to yield insignificant descendants at a finer scale, Shapiro developed the well-known embedded zerotree wavelet coder [16]. In contrast, large magnitude wavelet coefficients produced by true signal at finer scales are more likely to have significant parents at coarser scales. However, the coefficients caused by noise would decay rapidly along scales. With this observation, Xu et al. [17] multiplied the adjacent wavelet scales to sharpen edge structures while weakening noise. Sadler et al. [26] analyzed the multiscale wavelet products and applied them to step detection and estimation.

Another popular wavelet inter-scale dependency representation is through autoregressive models or hidden Markov models. By structuring the image wavelet coefficients as a quadtree along scales, Banham [21] imposed an autoregressive model that evolves from coarse scales to fine scales on the quadtree nodes. He then applied a multiscale Kalman smoothing filter after some prefiltering. The hidden Markov models (HMMs), especially the hidden Markov tree model (HMT) proposed by Crouse [18], characterize well the joint statistics of wavelet coefficients across scales. Each coefficient is assigned a hidden state, conditioned on which the coefficients are i.i.d Gaussian. HMT was further improved by Fan [19] with a new four-state model called HMT-2. Romberg [20] also simplified the HMT by exploiting the inherent self-similarity of real-world images.

In this section we present a new inter-scale dependencies exploited model based on the LMMSE algorithm. In the next section the model will be extended to inter-scale and intra-scale combined form. By implementing OWE on

the noisy image, we group the points with the same spatial position at J scales as a vector:

$$\vec{\boldsymbol{Z}}(m,n) = \mathrm{col}\{Z_1(m,n),\ldots,Z_J(m,n)\}, \tag{3.11}$$

where col denotes column vector. Also

$$\vec{\boldsymbol{Z}} = \vec{\boldsymbol{X}} + \vec{\boldsymbol{V}} \tag{3.12}$$

with

$$\vec{\boldsymbol{X}}(m,n) = \mathrm{col}\{X_1(m,n),\ldots,X_J(m,n)\} \quad \text{and}$$

$$\vec{\boldsymbol{V}}(m,n) = \mathrm{col}\{V_1(m,n),\ldots,V_J(m,n)\}, \tag{3.13}$$

where $\vec{\boldsymbol{v}}$ is the *Gaussian* vector noise and independent of $\vec{\boldsymbol{x}}$ The LMMSE of $\vec{\boldsymbol{x}}$ can be computed as

$$\hat{\vec{\boldsymbol{x}}} = P(P+R)^{-1}\vec{z}, \tag{3.14}$$

where P and R are the autocovariance matrices of $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{v}}$, respectively,

$$P = E[\vec{\boldsymbol{x}}\vec{\boldsymbol{x}}^{\mathrm{T}}] = E\begin{bmatrix} x_1^2 & \cdots & x_1x_J \\ \vdots & \ddots & \vdots \\ x_Jx_1 & \cdots & x_J^2 \end{bmatrix} \quad \text{and}$$

$$R = E[\vec{\boldsymbol{v}}\vec{\boldsymbol{v}}^{\mathrm{T}}] = E\begin{bmatrix} v_1^2 & \cdots & v_1v_J \\ \vdots & \ddots & \vdots \\ v_Jv_1 & \cdots & v_J^2 \end{bmatrix}. \tag{3.15}$$

The diagonal element of R, $E[v_j^2]$, is equal to σ_j^2, which can be computed by Eq. (3.5). Noise variables v_i and v_j, where $i \neq j$, are the projections of v on the wavelet bases ψ_i and ψ_j. They are correlated due to the similarity between ψ_i and ψ_j. The correlation coefficient is

$$\rho_{i,j} = \frac{\iint \psi_i(x,y)\psi_j(x,y)\,\mathrm{d}x\,\mathrm{d}y}{\sqrt{\iint \psi_i^2(x,y)\,\mathrm{d}x\,\mathrm{d}y}\sqrt{\iint \psi_j^2(x,y)\,\mathrm{d}x\,\mathrm{d}y}}, \tag{3.16}$$

where v_i and v_j are *jointly Gaussian* and their density is

$$p(v_i,v_j) = \frac{1}{2\pi\sigma_i\sigma_j\sqrt{1-\rho_{i,j}^2}}\mathrm{e}^{1/2(1-\rho_{i,j}^2)[v_i^2/\sigma_i^2 - 2\rho_{i,j}v_iv_j/\sigma_i\sigma_j + v_j^2/\sigma_j^2]}. \tag{3.17}$$

Thus the expectation $E[v_iv_j]$ is

$$R_{i,j} = E[v_iv_j] = \iint v_iv_j\,p(v_i,v_j)\,\mathrm{d}v_i\,\mathrm{d}v_j = \rho_{i,j}\sigma_i\sigma_j. \tag{3.18}$$

Each of the components of matrix P can be computed by

$$P_{i,j} = E[x_ix_j] = E[z_iz_j] - R_{i,j}, \tag{3.19}$$

where $E[z_iz_j]$ is estimated as

$$E[z_iz_j] \approx \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} Z_i(m,n)\cdot Z_j(m,n). \tag{3.20}$$

3.4. The intra- and inter-scale dependencies combined scheme

The inter-scale model proposed in Section 3.3 can be improved to merge intra-scale dependencies. This can be accomplished by estimating the covariance matrix P locally with a proper window.

A centered square window Λ is set for image $\vec{\boldsymbol{Z}}$. All the vector data in Λ is assumed to have the same statistics. The elements of local autocovariance matrix P^{Λ} are estimated by

$$\hat{P}_{i,j}^{\Lambda} \approx \max\left(0, \frac{1}{N_{\Lambda}}\sum_{(m,n)\in\Lambda} Z_i(m,n)Z_j(m,n) - R_{i,j}\right), \tag{3.21}$$

where N_{Λ} is the number of vector data in window Λ. Then the image coefficients within local window Λ are estimated by

$$\hat{\vec{\boldsymbol{x}}} = \hat{P}^{\Lambda}(\hat{P}^{\Lambda}+R)^{-1}\vec{z}. \tag{3.22}$$

The above LMMSE (3.22) incorporates the wavelet inter- and intra-scale dependency information together to smooth the noisy image coefficients.

3.5. The LMMSE and thresholding hybrid algorithm

Suppose the input image is decomposed into J wavelet scales, and it is found that setting $\vec{\boldsymbol{Z}}$ as a J-D vector by Eq. (3.11) will not yield satisfying results. This could be explained in two ways. Firstly, scale j is strongly correlated with scale $j+1$ but its correlation with scale $j+2, j+3,\ldots,J$ decreases rapidly. These coarser scales will not convey much useful information to improve the estimation of scale j. Secondly, a significant structure has larger spatial supports at coarser scales than finer scales. One point at a coarse scale may appear as an edge, but at finer scales the points with the same spatial orientation may be noise predominated. Thus fine scales may impose negative effects on coarse scales.

With these considerations, we will not employ fine scale measurements to update coarse scale image estimations. We recover X_j only from the measurements at adjacent scales j and $j+1$. Define the vector image $\vec{\boldsymbol{Z}}_j$ as

$$\vec{\boldsymbol{Z}}_j(m,n) = \mathrm{col}\{Z_j(m,n), Z_{j+1}(m,n)\}. \tag{3.23}$$

So

$$\vec{\boldsymbol{Z}}_j = \vec{\boldsymbol{X}}_j + \vec{\boldsymbol{V}}_j \tag{3.24}$$

with

$$\vec{\boldsymbol{X}}_j(m,n) = \mathrm{col}\{X_j(m,n), X_{j+1}(m,n)\} \quad \text{and}$$

$$\vec{\boldsymbol{V}}_j(m,n) = \mathrm{col}\{V_j(m,n), V_{j+1}(m,n)\}. \tag{3.25}$$

Applying the hybrid intra- and inter-scale LMMSE scheme in 3.4 to $\vec{\boldsymbol{Z}}_j$ the estimated image $\hat{\vec{\boldsymbol{x}}}_j$ is obtained. We preserve only the component $\hat{X}_j$ of it. Similarly, the estimation result $\hat{X}_{j+1}$ would be obtained by applying the LMMSE to data $\vec{\boldsymbol{Z}}_{j+1}$.

In the threshold-based denoising schemes, wavelet coefficients whose magnitude is below the preset threshold will be set to 0. This insignificant data is predominantly noise and should be discarded. To merge the merits of thresholding into the LMMSE scheme, we apply LMMSE only to the coefficients above a threshold and shrink those coefficients below the threshold to 0. The scale adaptive threshold employed in this paper is

$$t_j = c\sigma_j \tag{3.26}$$

with c being a constant between 3 and 3.5. Most of the Gaussian noise will fall into three times its standard deviation. This threshold was proposed by Pan et al. [19] and produces good results.

The LMMSE and thresholding hybrid algorithm is summarized by pseudocode as follows.

First implement the overcomplete wavelet expansion to get $Z_j,\ j = 1,2,\ldots,J$.
Loop for each scale j
{
  Form the vector image $\vec{\boldsymbol{Z}}_j$ *and compute the noise autocovariance matrix* R.
  Set threshold $t_j = c\sigma_j$ *and the centered square window* Λ_j *with a proper size.*
  Loop for each image block shaped by Λ_j
  {
  Estimate the autocovariance matrix $\hat{P}^{\Lambda_j}$.
  Loop for each coefficient within the block
   {
    If $|Z_j(m,n)| < t_j$
      $\hat{\boldsymbol{X}}_j(m,n) = 0$
    Else
      $\hat{\vec{\boldsymbol{x}}}_j(m,n) = \hat{P}^{\Lambda_j}(\hat{P}^{\Lambda_j} + R)^{-1}\hat{Z}_j(m,n)$
    End
  The first element of $\hat{\vec{\boldsymbol{x}}}_j(m,n), \hat{X}_j(m,n)$, *is preserved.*
   } *End of loop*
  } *End of loop*
} *End of loop*
Reconstruct the estimated image by $\hat{\boldsymbol{X}}_j,\ j = 1,2,\ldots,J$.

3.6. Wavelet bases selection

In Ref. [13], Liu and Moulin measured the wavelet coefficient dependencies by *mutual information*, which relates fundamentally to data compression, estimation and classification performance. They found that the *mutual information* depends on not only the used statistical models but also the employed wavelet filters. The denoising performance of the proposed scheme is also wavelet dependent. In this paper, we focus on the widely used compactly supported orthogonal and biorthogonal wavelets constructed by Daubechies [3].

We denote Daubechies' orthogonal wavelet as $Dau(N)$, where $N = 1,2,\ldots,\infty$ is the vanishing moment of the wavelet whose filter length is $2N$. The biorthogonal wavelet is denoted by $CDF(N,N')$, where N is the vanishing moment of the analytic wavelet filter and N' is that of the synthetic wavelet filter. Orthogonal wavelets $Dau(N)$, with the exception of the Haar wavelet, lack (anti-)symmetry. This is an important property in signal and image processing. Biorthogonal wavelets $CDF(N,N')$ discard the orthogonality to preserve the (anti-)symmetric property.

The proposed inter-scale model exploits the high correlation of adjacent wavelet scales. The denoising efficiency relies on how much positive information the coarser scale can convey to its adjacent finer scale. If a significant edge point occurs at a finer scale then it is expected that a corresponding edge point will appear with the same sign at coarser scale. Otherwise, a coarser scale may be detrimental to a finer scale estimation.

Define variable y_j as the product of two adjacent scale wavelet coefficients

$$y_j = x_j x_{j+1}. \tag{3.27}$$

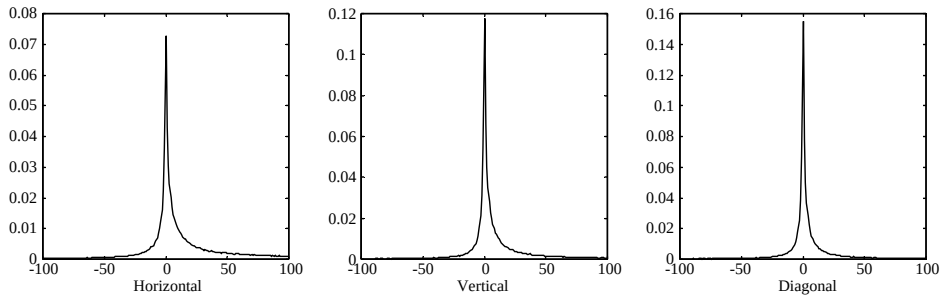
Different wavelet filters result in different distributions of y_j. The histograms of y_1 for image *Lena* obtained from four wavelets: *Dau* (2), *Dau* (4), *CDF* (1,3) and *CDF* (2,4) are illustrated in Fig. 4. (The histograms of y_j, $j > 1$, are similar to those of y_1 in shape.)

The expectation is that the histograms of y_j are asymmetrical and have heavy tails in the positive $\boldsymbol{X}$-axis, implying high correlation of adjacent scales. From Fig. 4 we can see that the histograms of y_1 become more symmetrical when the vanishing moments (as well as the length) of the wavelet filters increase. For example, the histograms for *Dau* (4) are more symmetrical than those for *Dau* (2). The same to *CDF* (1,3) and *CDF* (2,4). It is observed that the histograms' shape for biorthogonal wavelet *CDF* (1,3) is better than other wavelets. The positive tails are rather considerable. *CDF* (1,3) has only 1 vanishing moment for analytic filter while having 3 vanishing moments for synthetic filter to smooth noise. It is likely to be the most suitable wavelet for our algorithm.

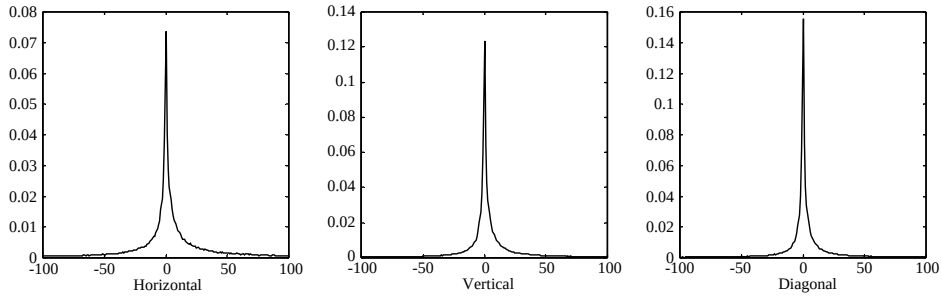
4. Experiments

We call the proposed algorithm in Section 3.6, M3. It is compared with Mıhçak's [14] intra-scale dependencies exploited scheme (M1) and the inter-scale dependencies exploited scheme (M2) presented in Section 3.3. It should be noted that we also hybridize scheme M2 with a similar thresholding strategy as in scheme M3. The wavelets *Dau* (2), *Dau* (4), *CDF* (1,3) and *CDF* (2,4) are employed in the experiments.

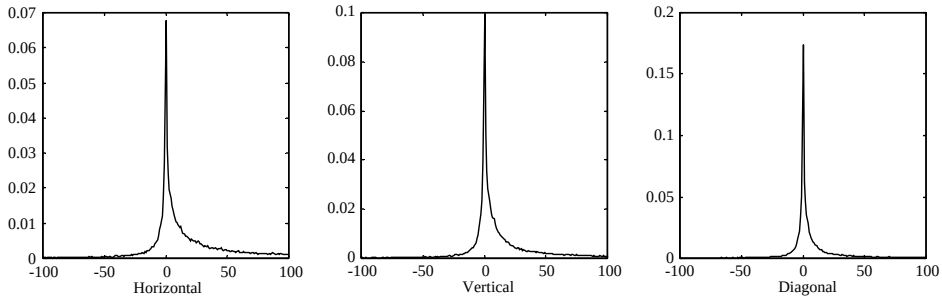
Several benchmark images have been used to test the three schemes and the conclusions are identical. The experimental results on images *Lena* and *Peppers* are reported here. The centered square window used in M1 and M3 is of 9×9 size. Three levels of Gaussian white noise with $\sigma = 20, 25, 30$ are added to *Lena* and *Peppers*, respectively. The denoising



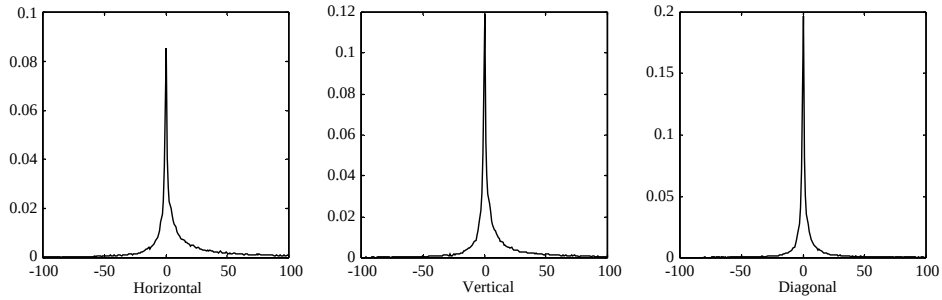
(a) *Dau* (2)



(b) *Dau* (4)



(c) *CDF* (1,3)



(d) *CDF* (2,4)

Fig. 4. Histograms of the products of the first two scale wavelet coefficients for different wavelets. (a) *Dau* (2); (b) *Dau* (4); (c) *CDF* (1, 3); (d) *CDF* (2, 4).

results by the three methods and four wavelets are listed in Tables 1 and 2.

It is observed that algorithm M3 always performs better than M1 and M2 for a fixed wavelet base. Furthermore, scheme M3 by wavelet *CDF* (1, 3) yields the best denoising performance. This is highlighted in Tables 1 and 2. While being a good bridge to exploit wavelet inter-scale dependencies, *CDF* (1, 3) is not so good at characterizing the wavelet

Table 1
SNR (dB) results for the three methods for image *Lena* in different noise levels

Lena		*Dau* (2)	*Dau* (4)	*CDF* (1,3)	*CDF* (2,4)
$\sigma = 20$	M1	25.74	25.69	25.40	25.66
	M2	25.76	25.70	25.83	25.35
	M3	25.95	25.82	26.09	25.67
$\sigma = 25$	M1	24.56	24.45	24.26	24.44
	M2	24.53	24.46	24.69	24.29
	M3	24.98	24.84	25.11	24.69
$\sigma = 30$	M1	23.53	23.39	23.28	23.37
	M2	23.55	23.46	23.71	23.27
	M3	24.09	23.94	24.21	23.81

Table 2
SNR (dB) results for the three methods with image *Peppers* for different noise levels

Peppers		*Dau* (2)	*Dau* (4)	*CDF* (1,3)	*CDF* (2,4)
$\sigma = 20$	M1	24.49	24.33	24.20	24.39
	M2	24.70	24.51	24.75	24.27
	M3	24.94	24.63	25.20	24.52
$\sigma = 25$	M1	23.42	23.24	23.17	23.31
	M2	23.74	23.55	23.82	23.29
	M3	24.09	23.76	24.32	23.62
$\sigma = 30$	M1	22.44	22.21	22.19	22.27
	M2	22.75	22.54	22.89	22.31
	M3	23.31	22.98	23.55	22.82

intra-scale dependencies and it is not an appropriate candidate for scheme M1. From Tables 1 and 2 we see that scheme M1 by *CDF* (1,3) always reports the poorest results. In contrast methods M2 and M3 take advantage of the wavelet inter-scale dependencies. So wavelets *CDF* (1,3) and *Dau* (2), which have stronger correlation between adjacent scales, are more suitable for the two schemes. In Fig. 5, the noisy *Lena* ($\sigma = 30$) and the three denoised versions by biorthogonal wavelet *CDF* (1,3) are illustrated. Fig. 6 shows the noisy *Peppers* ($\sigma = 30$) and the denoised images by *CDF* (1,3).

5. Conclusion

In this paper, we presented a wavelet based noisy image restoration approach that exploits both the intra- and inter-scale dependencies. The overcomplete wavelet expansion, which is more efficient in denoising than the orthogonal wavelet expansion, is employed in the scheme. The wavelet coefficients with the same spatial orientation at adjacent scales are combined as a 2-D vector. By applying the LMMSE to the vector variable, the inter-scale correlation information is utilized to update the estimation of one scale. To exploit the wavelet intra-scale dependencies, we used a centered square window to estimate the autocovariance matrices of wavelet coefficients locally. Finally, we combined the composite dependencies exploited algorithm with a thresholding strategy. The experimental results on benchmark images illustrate the proposed scheme outperforms the methods that only exploit the intrascale or interscale dependencies.

The performance of the proposed hybrid intrascale and interscale denoising scheme also depends on the selection of the wavelet filters. Four orthogonal and biorthogonal compactly supported wavelets were considered in this paper. The biorthogonal wavelet *CDF* (1,3), which is good at exploiting inter-scale dependencies, produces the best performance. The inter-scale correlation decreases with the increasing of the vanishing moment of the wavelet filter. The wavelet filter *CDF* (1,3) has the least vanishing moment for analytic filter (1-order) while having 3-order vanishing moments for the synthetic filter to smooth noise. The proposed scheme by *CDF* (1,3) produced good results in our experiments.

Fig. 5. Denoising results of *Lena* by wavelet *CDF* (1,3). (a) Noisy *Lena* ($\sigma = 30$). (b) Restored *Lena* by scheme M1 (SNR = 23.28 dB). (c) Restored *Lena* by scheme M2 (SNR = 23.71 dB). (d) Restored *Lena* by scheme M3 (SNR = 24.21 dB).

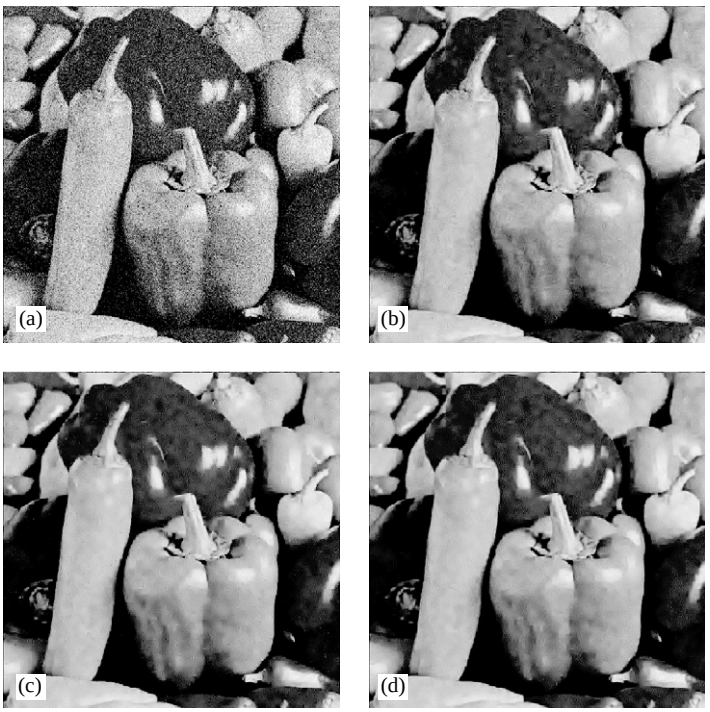


Fig. 6. Denoising results of *Peppers* by wavelet *CDF* (1,3). (a) Noisy *Peppers* ($\sigma = 30$). (b) Restored *Peppers* by scheme M1 (SNR = 22.19 dB). (c) Restored *Peppers* by scheme M2 (SNR = 22.89 dB). (d) Restored *Peppers* by scheme M3 (SNR = 23.55 dB).

References

[1] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693.

[2] S. Mallat, S. Zhong, Characterization of signals from multiscale edges, IEEE Trans. Pattern Anal. Mach. Intell. 14 (7) (1992) 710–732.

[3] I. Daubechies, Ten lectures on Wavelets, SIAM, Philadelphia, PA, 1992.

[4] M. Vetterli, C. Herley, Wavelet and filter banks: theory and design, IEEE Trans. Signal Process. 40 (9) (1992) 2207–2232.

[5] S.G. Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression, IEEE Trans. Image Process. 9 (9) (2000) 1532–1546.

[6] S.G. Chang, B. Yu, M. Vetterli, Spatially adaptive wavelet thresholding with context modeling for image denoising, IEEE Trans. Image Process. 9 (9) (2000) 1522–1531.

[7] D.L. Donoho, De-noising by soft thresholding, IEEE Trans. Inf. Theory 41 (5) (1995) 613–627.

[8] D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, J. Am. Stat. Assoc. 90 (12) (1995) 1200–1224.

[9] Q. Pan, L. Zhang, et al., Two denoising methods by wavelet transform, IEEE Trans. Signal Process. 47 (12) (1999) 3401–3406.

[10] F. Abramovich, T. Sapatinas, et al., Wavelet thresholding via a Bayesian approach, J. R. Stat. Soc. Ser. B 60 (1998) 725–749.

[11] A. Chambotle, R.A. Devore, et al., Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage, IEEE Trans. Image Process. 7 (7) (1998) 319–335.

[12] M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, Image coding using wavelet transform, IEEE Trans. Image Process. 1 (4) (1992) 205–220.

[13] J. Liu, P. Moulin, Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients, IEEE Trans. Image Process. 10 (11) (2001) 1647–1658.

[14] M.K. Mıhçak, I. Kozintsev, K. Ramchandran, P. Moulin, Low-complexity image denoising based on statistical modeling of wavelet coefficients, IEEE Signal Process. Lett. 6 (12) (1999) 300–303.

[15] S.D. Servetto, K. Ramchandran, M.T. Orchard, Image coding based on a morphological representation of wavelet data, IEEE Trans. Image Process. 8 (9) (1999) 1161–1174.

[16] J.M. Shapiro, Embedded image coding using zerostrees of wavelet coefficients, IEEE Trans. Signal Process. 41 (12) (1993) 3445–3462.

[17] Y. Xu, et al., Wavelet transform domain filters: a spatially selective noise filtration technique, IEEE Trans. Image Process. 3 (11) (1994) 747–758.

[18] M. Crouse, R. Nowak, R. Baraniuk, Wavelet-based statistical signal processing using hidden Markov models, IEEE Trans. Signal Process. 42 (4) (1998) 886–902.

[19] G. Fan, X.G. Xia, Improved hidden Markov models in the wavelet-domain, IEEE Trans. Signal Process. 49 (1) (2001) 115–120.

[20] J.K. Romberg, H. Choi, R.G. Baraniuk, Bayesian tree-structured image modeling using wavelet-domain hidden Markov models, IEEE Trans. Image Process. 10 (7) (2001) 1056–1068.

[21] M.R. Banham, A.K. Katsaggelos, Spatially adaptive wavelet-based multiscale image restoration, IEEE Trans. Image Process. 5 (4) (1996) 619–634.

[22] J. Portilla, V. Strela, M.J. Wainwright, E.P. Simoncelli, Adaptive wiener denoising using a Gaussian scale mixture model in the wavelet domain, Proceedings of the Eight International Conference on Image Processing, Thessaloniki, Greece, 2001, pp. II. 37–40.

[23] J. Liu, P. Moulin, Image denoising based on scale-space mixture modeling for wavelet coefficients, Proceedings of the International Conference on Image Processing (ICIP'99) Kobe, Japan, 1999, pp. I. 386–390.

[24] R.R. Coifman, D.L. Donoho, Translation-invariant de-noising, in: A. Antoniadis, G. Oppenheim (Eds.), Wavelet and Statistics, Springer, Berlin, Germany, 1995.

[25] E.W. Karmen, J.K. Su, Introduction to Optimal Estimation, Springer, London, 1999.

[26] B.M. Sadler, A. Swami, Analysis of multiscale products for step detection and estimation, IEEE Trans. Inf. Theory 45 (4) (1999) 1043–1051.

About the Author—LEI ZHANG was born in 1974 in P.R. China. He received the M.Sc and Ph.D Degree in Electrical and Engineering from Northwestern Polytechnical University, Xi'an, P.R. China, respectively, in 1998 and 2001. He is currently working as a research associate in the Dept. of Computing, The Hong Kong Polytechnic University. His research interests include digital signal and image processing, wavelet transform and optimal estimation theory.

About the Author—PAUL BAO received the B.Sc (82) and M.Sc (84) Degrees in Mathematics and Computer Science from Chongqing University and Jiao Tong University respectively and Ph.D. (88) in Computer Science from University of Calgary, Canada for the work on computer graphics, computational geometry and algorithm design. He served on the faculty of Computer Science Department, University of Calgary from 1988–1990 and then proceeded to work in IBM Canada as a staff analyst in the areas of Object-Oriented Technologies and Distributed Computing Systems from 1990–1995. Since July 1995, he has been with Computing Department, the Hong Kong Polytechnic University where he is now an associate professor. Dr. Bao's research interests included Computer Graphics, Image-based Rendering, Image Compression, Neural Network for Image Processing, Compression and fuzzy-based Financial Engineering.

About the Author—XIAOLIN WU received his B.Sc. in computer science from Wuhan Cehui Technical University, Wuhan, China, in 1982 and Ph.D. in computer science from University of Calgary, Calgary, Canada, in 1988. He is currently a professor with Department of Computer Science, University of Western Ontario, London, Canada. His research interests include algorithms for visual computing and communication, data compression, image processing and computer graphics. He has published more than 70 research papers in those areas. Professor Wu is an active JPEG member and the 1998 UWO Distinguished Research Professor.