Spatially Adaptive Wavelet-Based Multiscale Image Restoration

Mark R. Banham, Member, IEEE, and Aggelos K. Katsaggelos, Senior Member, IEEE

Abstract—In this paper, we present a new spatially adaptive approach to the restoration of noisy blurred images, which is particularly effective at producing sharp deconvolution while suppressing the noise in the flat regions of an image. This is accomplished through a multiscale Kalman smoothing filter applied to a prefiltered observed image in the discrete, separable, 2-D wavelet domain. The prefiltering step involves constrained least-squares filtering based on optimal choices for the regularization parameter. This leads to a reduction in the support of the required state vectors of the multiscale restoration filter in the wavelet domain and improvement in the computational efficiency of the multiscale filter. The proposed method has the benefit that the majority of the regularization, or noise suppression, of the restoration is accomplished by the efficient multiscale filtering of wavelet detail coefficients ordered on quadtrees. Not only does this lead to potential parallel implementation schemes, but it permits adaptivity to the local edge information in the image. In particular, this method changes filter parameters depending on scale, local signal-to-noise ratio (SNR), and orientation. Because the wavelet detail coefficients are a manifestation of the multiscale edge information in an image, this algorithm may be viewed as an “edge-adaptive” multiscale restoration approach.

I. INTRODUCTION

There has been a significant trend in recent research regarding the use of multiscale approaches to image and signal processing. The increased attention to this field is largely motivated by great activity in the field of wavelets. The use of wavelets has become quite extensive in the areas of mathematics and mathematical physics [1], numerical analysis [2], and single and multidimensional signal processing [3]. As seen by this research, many processes occurring in nature may be modeled effectively in some multiscale context. Images, in particular, lend themselves to this type of modeling because of the existence of structures and details that may be characterized at different resolutions within the image. This is especially evident when considering the presence of an object with texture information inside it. The outline of the object may be defined by a very coarse representation of the edges, but the texture information inside the object may occur at a high frequency, and could require a fine-scale, high-resolution representation. The edges in an image hold a great deal of the perceptual information, as they define the various regions of interest in an image, and thus it is appropriate to gear image processing algorithms around these edges. In the field of image processing, it is useful to view the wavelet transform as a tool for manipulating and understanding the behavior of edges in an image. This perspective leads to algorithms in coding [4], reconstruction [5], and enhancement [6]–[8], which are sensitive to the perceptual importance of edges to the human viewer.

At the same time that interest has flourished in wavelet-based techniques, a very important related theory describing optimal estimation algorithms for signals that appear in a multiscale framework has been developed [9], [10]. These optimal estimation ideas have led to work in nonstationary signal parameter estimation [11], motion estimation [12], texture modeling [13], and data fusion [14]. The statistical theory of multiscale stochastic processes can lead to an optimal way to view the behavior of wavelet scaling or detail coefficients as they evolve in scale. Utilizing multiscale stochastic estimation theory, we formalize an adaptive restoration approach here, which consistently produces perceptually pleasing results due to the high-quality preservation of edges in the restored images.

By way of definition, let a multiscale signal represent any collection of discrete data that appears at different scales (fine to coarse) whose support, or sampling rate, is dependent on scale. Here, we will consider the scale and the sampling rate of the data to decrease as data becomes coarser in this representation. The multiscale relationship between scaling coefficients of the wavelet transform is easily modeled as a weighted averaging process as the scale decreases. Because of this, multisresolution estimation approaches have generally concentrated on processing the successive low resolution approximations of an image, which are implicit in a wavelet decomposition [15], or on processing a quadtree decomposition of the image intensity data [16]. It would, however, also be beneficial to apply a multiscale filter directly to the wavelet detail coefficients of an image in order to take advantage of the edge characteristics of this information at different orientations. This leads to the need for a valid description of the relationships between detail coefficients at different scales, which was investigated in [17] and [18]. To simplify the discussion for the remainder of this paper, we hereafter refer
to the detail coefficients of the wavelet transform as simply the "wavelet coefficients," and the low resolution approximation coefficients as simply the "scaling coefficients," as is the convention.

The use of wavelet coefficients in the area of noise suppression and restoration has been examined in several studies. One approach has considered applying maximum a posteriori (MAP) estimation to each wavelet coefficient of a noisy image to obtain a smoother image [19]. This approach also considered a hierarchical prediction of wavelet coefficients; however, the à trous wavelet algorithm [20], which includes no downsampling, was used, thus making the prediction operation quite simple. The idea of wavelet shrinkage [21] is also based on a statistical model of the wavelet coefficients and considers applying a soft threshold to each noisy wavelet coefficient to reduce the noise. In situations where the degradation contains a blurring operation, statistical noise smoothing in the wavelet domain has been applied to the residual at each step of an iterative Richardson–Lucy image restoration algorithm in [22], and as a prior step to iterative deconvolution in [19].

These ideas have primarily focused on processing only noisy wavelet coefficients directly. In this work, we incorporate the processing of noisy and blurred data in a more formal framework directly in the wavelet domain. This is accomplished by first casting the classical image restoration problem into the wavelet domain. We then consider the behavior of the blur operator in the wavelet domain, and examine the length of the state vectors containing wavelet coefficients in an optimal state space multiscale Kalman smoothing filter. The use of a multiscale Kalman filter applied directly to the wavelet coefficients is a large part of the novelty found in this work. By prefiltering the data in order to reduce the length of each state vector, we eliminate the need for large state vectors containing information from different scales. This also leads to a very efficient multiscale filter, which may be implemented in scalar form. It will be shown that this approach produces good results in comparison to standard direct restoration approaches, and spatially adaptive restoration based on iterative methods. In the next section, we introduce the image restoration problem and discuss some of the relevant classical methods for obtaining a restored image. In Section III, we show the optimal recursive multiscale formulation of the restoration problem in the wavelet domain, and Section IV introduces a reduced order model obtained by prefiltering this wavelet data. Section V shows a comparison of the results obtained by this adaptive method and the classical linear techniques are discussed here. Section VI presents some conclusions.

II. THE IMAGE RESTORATION PROBLEM

The general model for a degraded image may be written as

\[ y = Df + n \]  \hspace{1cm} (1)

where a lexicographic ordering of the original image, \( f \), the observed image, \( y \), and the observation noise, \( n \), is used. The observation noise will be treated as additive white Gaussian noise in all of the examples examined in this paper.

Furthermore, the blur operator will be treated as known a priori here. In (1), \( D \) usually represents a block-circulant degradation operator. There are numerous approaches to solving this problem, including direct, iterative, and recursive methods [23]. Here, we are interested in taking advantage of the efficiency of direct methods while combining the adaptivity available in multiscale recursive approaches, with minimal added complexity.

In order to compare the spatially adaptive approach introduced in this paper to standard direct methods, some background is provided next. Classical direct approaches for solving this ill-posed problem have relied on finding an \( \hat{f} \) to satisfy

\[ \hat{f} = \arg \left\{ \min_{f} \| y - Df \|^2 \right\} \]  \hspace{1cm} (2)

subject to some constraints based on our prior knowledge about the image [24]. In the stochastic formulation, this prior knowledge appears in the form of the power spectra of the additive noise and the image. In the deterministic formulation, this prior information is manifested in a smoothness constraint on the restored image. In either case, (2) leads to a solution of the form

\[ (D^T D + M)\hat{f} = D^T y \]  \hspace{1cm} (3)

where \( M \) represents the prior constraint information. Given a circulant assumption for all of the matrices involved, (3) is easily solved in the discrete frequency domain. In the stochastic case, the choice of \( M = B_{nn} R_{ff}^{-1} \), where \( R_{ff} \) represents the covariance matrix of the original image and \( B_{nn} \) represents the covariance matrix of the noise, leads to the Wiener filter. This approach relies on the availability of a good statistical model of the image and of the noise. In the case where the blurring component of the degradation is consequential, we often have severe difficulty in obtaining a good estimate of the image power spectrum from the degraded data. In particular, the high-frequency components of this spectrum are often lost or inaccurately estimated. If the noise variance estimate is biased as well, we often end up with a restored image that is fairly smooth. While this leads to good improvement in objective metrics such as improvement in mean squared error, it does not necessarily produce the best restoration from a perceptual viewpoint. This will be clear in the examples shown in Section V.

In deterministic restoration, it is possible to choose \( M = \alpha C^T C \), where \( C \) is taken to be a differential operator such as the 2-D discrete Laplacian, and \( \alpha \) is a regularization parameter chosen to govern the trade-off between the smoothness of the restored image and the fit of the solution to (2). The choice of this parameter can be quite complex, and is the subject of many articles in the area of constrained least squares (CLS) restoration, for example in [25]–[27].

If this parameter is chosen to be very small, the restored image is often quite sharp, but subject to severe noise amplification in the high-frequency portions of the spectrum. From an objective viewpoint, the improvement with this approach will often be small, but the results may be perceptually more pleasing than a smoother
solution provided by the Wiener filter. This will also become apparent with the results shown in Section V.

Based on an understanding of direct approaches, it is possible to formulate an adaptive technique that has the benefits of both the smooth stochastic solution and the noisy deterministic solution. As mentioned, the direct methods discussed are usually implemented in the discrete frequency domain, and are among the most effective and computationally simple methods used in image restoration. Older methods for spatially varying restoration have included sectioning the image and applying fixed models to each section of the image. This type of approach may be limited in its usability because of the significant \textit{a priori} information needed in making the required segmentation. In [25] and [28], for example, a different iterative restoration filter was used at each picture element based on a measure of the local spatial activity, which determines the noise visibility function. In [25] and [29], the image was segmented into a number of regions also based on a measure of the spatial activity, and a different iterative restoration filter was applied to each region.

Various ways for obtaining accurate measures of the spatial activity were presented and investigated in [30]. Since the measure of the spatial activity, or equivalently the segmentation, is based on the noisy and blurred image, artifacts may be introduced in the restored image. To alleviate this problem, in [31] the spatial activity or segmentation information was updated at each iteration step based on the partially restored image. In the context of spatial Kalman filtering [32], the image is often modeled as locally stationary with autoregressive (AR) models switching throughout the image. Again, the appropriate models for each region of the image need to be estimated, making such approaches somewhat more complicated than direct methods.

In developing an adaptive technique to improve upon these approaches, it is necessary to consider minimizing the additional computational load needed to improve the restoration results over these classical techniques. One avenue that leads to computationally efficient algorithms is that of multiscale filtering. In the next section, we present the background and formulation of regularized multiscale filtering, and discuss its use in the context of the detail coefficients of the wavelet transform.

III. MULTISCALE FILTERING IN THE WAVELET DOMAIN

In the area of stochastic multiscale processes, much important work has appeared in recent years on the topic of estimation [9], [10], [13]. This work is largely dedicated to developing a calculus for working with multiscale processes occurring on dyadic trees and quadrees. Multiscale algorithms may be formulated to arrive at the optimal estimate for the value of a node on a tree, given observations at all of the other nodes on a tree within some neighborhood. Because we are interested in applying multiscale estimation to the image restoration problem, here we limit the discussion to multiscale stochastic processes defined on quadrees, as in [13].

Consider a multiscale process that has measurements available on all nodes of a quadtree of some specified number of levels. Fig. 1 shows an example of a quadtree of three levels.

The notation used to describe the relationships between nodes on these trees has been thoroughly defined in [10]. Here we make use of this notation, and refer the reader to [10] and [13] for extensive discussions of the notation of 1-D and 2-D multiscale processes. Given a multiscale process, we can impose an AR model that evolves from coarse to fine scales on this structure, according to

\[
z(s) = A(s)z(s^c) + B(s)e(s).
\]  
(4)

Here, \(z(s)\) represents a state vector associated with node \(s\), \(z(s^c)\) represents the state vector of its parent, \(e(s) \sim N(0, I)\), a normally distributed zero-mean random process with identity covariance, and \(A(s)\) and \(B(s)\) are matrices of appropriate order. The noise represented by the \(B(s)\) term accounts for error in the cross-scale prediction. This model assumes that the state of a node is dependent only on the state of the node lying above it on the tree. In general, the size of these vectors and matrices is problem-dependent, as will be shown. It is also possible to consider the case when these components collapse to scalars, a technique that will be employed later. For the moment, however, the general case is considered.

Before proceeding, it is important to understand what type of behavior this implies in the context of images. Consider that an image may be represented as a multiscale process by utilizing an orthogonal wavelet transform of the original data. If we view the nodes on our tree as the wavelet coefficients, then (4) defines a causal relationship between edges of the image at a coarse scale and those in the same spatial position at a finer scale. The notion of exploiting the relationship between coefficients at different scales has been explored before, particularly in the context of image coding in [4], [33], and [34].

A. The Wavelet Transform

In order to better understand how the wavelet transform leads to a multiscale image processing framework, we first review the important aspects of this decomposition. The discrete wavelet transform may be interpreted as a tool for decomposing a signal over an orthonormal basis whose basis functions are translates and dilates of a single analysis function, \(\psi(x)\), which is generally called the mother wavelet. An important feature of such a basis is that it provides a nonuniform partitioning of the time-frequency plane. This leads to processing techniques that can view lowpass structures at a coarse scale and higher resolution features at a finely sampled scale. The discrete wavelet transform produces a set of wavelet coefficients and a set of scaling coefficients (which are obtained by expanding the signal over the scaling function
The wavelet coefficients represent bandpass-filtered versions of the original signal, and the scaling coefficients represent lowpass-filtered versions.

The scaling function provides the means to generate a 2-D multiresolution analysis of an image [5], and thus this function is frequently chosen to provide good energy compaction of the image into the low-resolution subbands. In two dimensions, the dyadic wavelet transform requires that each 2-D low resolution approximation of a signal be a refinement of the previous low resolution approximation, which is indicated in discrete terms by the relationship

$$\phi(x,y) = \sum_{m,n} h(m,n) \phi(2x - m, 2y - n).$$  \hspace{1cm} (5)

This then leads to the choice of three wavelet functions $$\psi^\lambda(x,y)$$, for $$\lambda = 1, 2, 3$$, which must satisfy a similar condition [36], namely

$$\psi^\lambda(x,y) = \sum_{m,n} g^\lambda(m,n) \phi(2x - m, 2y - n).$$  \hspace{1cm} (6)

Given this type of dyadic multiresolution analysis, it is possible to pick the finite impulse response (FIR) filters $$h(m,n)$$ and $$g^\lambda(m,n)$$, for $$\lambda = 1, 2, 3$$ in order to satisfy the perfect reconstruction conditions that must apply. It is convenient from an implementation viewpoint to choose separable filters for the required FIR filters, which is the approach followed here. To compute the 2-D wavelet decomposition with these filters, the original discrete image $$f(m,n)$$ is decomposed according to

$$v_j(m,n) = \sum_{k,l \in S_j} h(2m - k, 2n - l) f(k,l)$$  \hspace{1cm} (7)

$$w_j^\lambda(m,n) = \sum_{k,l \in S_j} g^\lambda(2m - k, 2n - l) f(k,l)$$  \hspace{1cm} (8)

where $$v_j(m,n)$$ represents the scaling coefficients of $$f(m,n)$$, $$w_j^\lambda(m,n)$$ represents the wavelet coefficients of $$f(m,n)$$, and $$S_j$$ indicates the region of support of the low resolution approximation at Scale $$j$$. Then, each successive lowpass approximation may be further decomposed according to

$$v_{j-1}(m,n) = \sum_{k,l \in S_j} h(2m - k, 2n - l) v_j(k,l)$$  \hspace{1cm} (9)

$$w_{j-1}^\lambda(m,n) = \sum_{k,l \in S_j} g^\lambda(2m - k, 2n - l) v_j(k,l)$$  \hspace{1cm} (10)

giving a recursive algorithm for decomposing an image in wavelet space. The original image can be reconstructed from the low-resolution approximation at the coarsest scale and all of the wavelet coefficients at finer scales. For our purposes here, as mentioned, we limit the use of the wavelet transform to the separable case, where the wavelets used are 1-D functions, and the image is decomposed separately along its rows and columns. This leads to a fast and efficient discrete wavelet transform, which provides information about the edges in an image in the form of subbands along three different orientations: horizontal ($f_{HL}$), vertical ($f_{HL}$), and diagonal ($f_{HH}$) [35]. The subband ($f_{LL}$) is recursively decomposed in the same way.

B. Problem Formulation in the Wavelet Domain

We now investigate the behavior of (1) in the wavelet domain. As mentioned before, the goal of this formulation will be to develop an “edge-adaptive” restoration algorithm by utilizing the wavelet coefficients as the data that we restore. It has been shown in our previous work that one way to deal with (1) in the wavelet domain is to use an orthogonal matrix $$W$$ composed of the 2-D wavelet transform filter coefficients, and decompose the problem one level at a time into a multichannel problem. Accordingly, we can decompose one level of the problem as

$$W y = WD W^T W f + W n$$

$$\tilde{y} = \tilde{D} f + \tilde{n}.$$

Throughout this paper, we will use the notation to indicate that the signal is in the wavelet domain. The multichannel structure of this representation now leads to special circulant matrix assumptions for the operator $$D$$ in the wavelet domain (for a precise definition of $$W$$ and an explanation of the matrix formulation, the reader is referred to [37] and [38]. Given that we further decompose the problem into a multiscale framework by recursively decomposing the low resolution approximations, the blur operator will contain cross-scale as well as cross-orientation terms. The wavelet transform of the degraded image is, therefore, just the result of taking a different degradation operator and applying it directly to the wavelet transform of the original image. This simply means that each blurred wavelet coefficient is obtained by a weighted sum of wavelet coefficients from its own subband (at a given scale and orientation) and coefficients representing the same spatial neighborhood in other scales and orientations.

By writing the observation equation in terms of each wavelet coefficient, which we now call $$\tilde{y}(s)$$, where $$s$$ represents the position in the framework we will use, we get the following representation:

$$\tilde{y}(s) = \tilde{D}^T(s) \tilde{f}(s) + \tilde{n}(s).$$  \hspace{1cm} (12)

Here, $$\tilde{D}(s)$$ is a vector composed of the auto- and cross-subband blur terms contributing to $$\tilde{y}(s)$$, and $$\tilde{f}(s)$$ represents a vector of all wavelet and scaling coefficients of the original image that fall within the neighborhood of coefficients contributing to $$\tilde{y}(s)$$. $$\tilde{D}(s)$$ is a vector here, as opposed to the matrix $$\tilde{D}$$ formulation in (11), to reflect the fact that $$\tilde{y}(s)$$ is a scalar. Subsequently, $$\tilde{n}(s)$$ is also a scalar representing the additive noise in the wavelet domain at position $$s$$.

The degradation operator in the wavelet domain is obtained by filtering the coefficients of the blur operator and decimating to get the blur in each subband, and then refiltering to get the cross-blur terms between subbands. In the simple case of an $$N \times 1$$ blur, with an $$M$$-tap wavelet filter, one filtering pass followed by decimation would result in a signal of $$(N + M - 1)/2$$ coefficients. Consider next that (11) shows the 2-D wavelet transform applied in an orthogonal matrix both forward and backward onto the original blur operator $$D$$, and, thus, many such filtering and decimation operations are needed to obtain the final blur coefficients. For a blurred coefficient in the wavelet domain, then, the required length of
$\hat{D}(s)$ may be quite large, but always depends on the number of levels in the wavelet decomposition, and the extent of the wavelet filters and original blur operator.

Equation (12) provides us with a means to restore each wavelet coefficient and adapt the restoration process to the characteristics of the coefficient, such as its scale, orientation, and magnitude. The crucial problem with the representation of (12) is that it is dependent on “cross-scale” terms. We will see that this presents some difficulty in processing this data in a multiscale optimal estimation framework. However, by prefiltering the data to eliminate the cross-scale information, it is possible to apply such techniques, as we will discuss in Section IV. The basic approach to multiscale optimal estimation is presented next.

C. Multiscale Estimation of Wavelet Coefficients

Let $\hat{g}(s)$ represent a wavelet coefficient from a given orientation at Scale $j$, and $\hat{g}(s\beta_i)$ represent a corresponding wavelet coefficient at the next coarsest scale, $j – 1$. We will also use the notation that the corresponding children of a node at the next finest scale, $j + 1$, be represented by $\hat{g}(s\beta_i)$. In the 2-D wavelet transform case, $i = 1, \cdots, 4$. For the purposes of explanation, we will limit the number of scales in this decomposition to three. So, for each orientation, we will have three subbands containing wavelet coefficients at different scales. In order to filter this wavelet-based data, it is necessary to place these coefficients into some type of data structure. For each coefficient in the coarsest band at each orientation, there are four coefficients at the next scale and 16 at the finest scale. These 21 coefficients represent the edge information along one orientation for an $8 \times 8$ spatial region of the image. The relationships of these coefficients may be represented with a quadtree, and each tree may be independently filtered. Such a quadtree is shown in Fig. 2.

Consider (4), in which we defined a multiscale AR process. Casting the observed wavelet coefficients into the quadtree structures defined above, we can express the relationships of the vectors $\hat{f}(s)$ in (12) as state vectors associated with each node in a quadtree. This relationship can be written as

$$\hat{f}(s) = A(s)\hat{f}(s\gamma) + B(s)e(s).$$

In multiscale state space notation, (12) represents the observation at a given scale, and (13) represents the dynamic relationship between the state vectors at different scales. The model noise is included here to account for inaccuracies in the prediction component of the model provided by $A(s)$. Again, we refer to $\hat{f}(s)$ as a vector in this general formulation, whose size is dependent upon the problem being solved. That size is a function of the support of the blur operator, the number of levels of wavelet decomposition, and the length of the wavelet filters themselves. It is important to remember, however, that this vector only contains a small subset of the total wavelet coefficients of the original image (in fact, just those that are linearly combined to give a single blurred coefficient). We utilize this notation to recursively estimate the state at each node on a given quadtree based on the observed wavelet coefficients associated with all other nodes on that tree.

The development of a scale recursive filter to achieve such estimation has been handled in detail by Chou [9] in one dimension, and then by Luettgen [13] in two dimensions. One optimal algorithm for estimating the value of all nodes in a multiscale stochastic process can be seen as a Kalman smoothing filter, with an additional merging step included to account for the variation in signal support across scales. This filter sweeps from fine to coarse scales, and then from coarse to fine scales, incorporating all of the data on the tree into each estimate.

Outlined next are the filtering equations for this estimation procedure. This technique is based on a generalization of the Rauch–Tung–Striebel smoothing algorithm developed primarily in [9]. The upward sweep of the algorithm consists of a Kalman filtering pass giving an estimate of each node based on its descendant nodes in the tree. The downward sweep then computes the smoothed estimates [9], [10], [13].

The initialization takes place at the finest scale, by setting the values of the predicted states and error covariances accordingly. Given (13), the covariance of the state $P_s$ evolves according to

$$P_s = A(s)P_sA^T(s) + B(s)B^T(s).$$

Initializing the covariance of the state to zero at the coarsest scale, the value of this covariance at the finest scale may then be found according to (14). The predicted error covariance at node $s$ of the finest scale is thus chosen as

$$P(s|s+) = P_s$$

where the notation $s|s+$ indicates that the noisy data at the descendant nodes of $s$, but not at $s$ itself, are taken into account. The prediction of the state is

$$\tilde{f}(s|s+) = [0]$$

where [0] represents a vector of zeros of appropriate length. We can then write the Kalman filtering pass from fine to coarse scales.

**Upward Sweep**

$$\tilde{f}(s|s+) = \hat{f}(s|s+) + K(s)[\hat{g}(s) - \hat{D}(s)\tilde{f}(s|s+)$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$

$$K(s) = P(s|s+)\hat{D}(s)[\hat{D}(s)P(s|s+)\hat{D}(s)]^{-1},$$

$$P(s|s+) = [I - K(s)\hat{D}(s)]P(s|s+)[I - K(s)\hat{D}(s)]^T$$
where $R_{\tilde{\gamma}}(s) = E\{\tilde{\gamma}(s)\tilde{\gamma}(s)^T\}$. It is next necessary to define an upward model, related to (13), as
\[
\tilde{f}(s) = F(s)\tilde{f}(s) + \tilde{e}(s)
\] (20)
where
\[
F(s) = P_{\tilde{\gamma}}A(s)P_{\tilde{\gamma}}^{-1}.
\] (21)
This model assumes that $A(s)$ is invertible for all $s$; however, $A(s)$ and $F(s)$ need not, in general, be orthogonal. By denoting the four descendants of $f(s)$ by $f(s\beta_i), i = 1, \ldots, 4$, we have
\[
\tilde{f}(s) = F(s, s\beta_i)\tilde{f}(s\beta_i)
\] (22)
\[
P(s, s\beta_i) = F(s, s\beta_i)P(s\beta_i)F^T(s, s\beta_i) + Q(s, s\beta_i)
\] (23)
\[
Q(s) = E\{\tilde{e}(s)\tilde{e}(s)^T\}.
\] (24)

Now, in (17), the prediction $\tilde{f}(s) + \tilde{e}(s)$ is based on the contributions of each of the child nodes beneath $f(s)$, according to
\[
\tilde{f}(s) + \tilde{e}(s) = P(s)\tilde{f}(s) + \sum_{i=1}^{4} P^{-1}(s|s\beta_i)\tilde{f}(s\beta_i)
\] (25)
where
\[
P(s) = \left[-3P_{\tilde{\gamma}}^{-1} + \sum_{i=1}^{4} P^{-1}(s|s\beta_i)\right]^{-1}.
\] (26)
Equations (25) and (26) contain the merging operations that distinguish this multiscale filter from classical implementations of the Kalman filter. The prediction-update structure of the recursion is otherwise preserved, however.

At the top of the tree, the optimal estimates of the state of the root node and corresponding error covariance are
\[
\tilde{f}(0|0) = \tilde{f}(0|0)
\] (27)
and
\[
P(0) = P(0|0).
\] (28)

We now incorporate the estimates computed during the upward sweep to obtain the optimal estimate of each node on the tree.

\[
\tilde{f}(s) = \tilde{f}(s) + J(s)\left[\tilde{f}(s) - \tilde{f}(s|s)\right]
\] (29)
\[
P(s) = P(s|s) + J(s)P(s|s)J^T(s)
\] (30)
\[
J(s) = P(s|s)F^T(s)P^{-1}(s|s).\]
(31)
The smoothed optimal estimate of the state vector associated with each node on the tree is then given by $\tilde{f}(s)$. We should note that this filtering operation may be applied independently to each quadtree of wavelet coefficients, and thus leads to efficient parallel implementation schemes.

IV. ADAPTIVE RESTORATION ALGORITHM

One important point that is crucial in the use of multiscale estimation is the number of scales used in the estimation process. For most practical applications involving images, the number of scales used in a wavelet decomposition is small (usually three to five). This is a result of the finite support of the data, and the presence of extended structures in the image that become obscured when the scale is too coarse. Thus, the classical idea of prediction in the state space is hindered by the small amount of data in the direction of the recursion. When the state vectors at each scale are large, the prediction must also account for a large amount of data with a very limited history of accurate estimates. Depending on the accuracy of the filter model, it may be impossible to obtain good estimates for all values of the state vectors. As mentioned in Section III-B, the state vectors representing the problem in the wavelet domain are necessary quite large. The length of these vectors is a function of the length of the blur in the spatial domain, and the extent of the wavelet filters used. In addition, the fact that the estimates of restored coefficients are based on data that comes from different scales complicates the use of (13) because it implicitly assumes that each state vector is composed of estimates from a single scale. Therefore, a procedure that removes the cross-scale dependency within the state vectors will allow us to utilize the multiscale structure governed by (13), and also to substantially reduce the length of the state vectors at each scale, making the multiscale implementation more efficient.

A. Model Order Reduction by Prefiltering

In practical Kalman filtering theory it is often desirable to attenuate some states such that they can be safely ignored, thus simplifying the implementation of the filter and reducing the amount of data needing to be processed [39]. In the case of this restoration problem, we also wish to reduce the length of the state vectors at each scale, and, in effect, remove the dependency on cross-scale components within the vectors. This will lead to a more appropriate filter model for the multiscale case. It is possible to do this by applying a filter that eliminates the correlation between terms in the spatial support of each state vector. Looking again at (12), we see that the term $D(s)$ determines the necessary support of the state vector at node $s$. If the blur support in the wavelet domain can be reduced, i.e., $D(s) = 1$, we eliminate the cross-scale information in $f(s)$, and reduce the complexity of our filter. This can be accomplished with an appropriately chosen prefilter applied before the multiscale Kalman smoother, as in Fig. 3.

Consider, then, a new filter model that is governed by the equations
\[
\tilde{z}(s) = \tilde{f}(s) + \tilde{n}(s)
\] (32)
\[
\tilde{f}(s) = A(s)\tilde{f}(s) + B(s)e(s).
\] (33)
In (32), \( \tilde{z}(s) \) represents the original observation \( \tilde{y}(s) \) after prefiltering. Notice that (33) is similar to (13), the only difference being that the states now contain the wavelet coefficients of some prefiltered data. Thus, these points can now be modeled as scalars, and \( \tilde{f}_F(s), A(s), \) and \( B(s) \) are hereafter treated as such.

Since we wish to obtain an observation with \( \hat{D}(s) = 1 \), as in (32), our prefilter must perform a deconvolution operation. Based on the discussion in Section II, we would like to have this prefilter provide a solution that is very sharp, and yet possibly noisy. That is, an underregularized solution would be beneficial from the viewpoint of the multiscale noise smoothing filter. In order to achieve that, we can consider using a constrained least squares filter to perform this operation, according to

\[
\tilde{z} = F \hat{y}, \quad F = (\hat{D}^T \hat{D} + \alpha \hat{C}_r^T \hat{C}_r)^{-1} \hat{D}^T \tag{34}
\]

\[
F = (D^T D + \alpha C_r^T C_r)^{-1} D^T \tag{35}
\]

where we have written this prefiltering operation in terms of the lexicographically stacked wavelet domain data. The border problem in the restoration filter is treated by assuming that the original degradation operator was applied with circular convolution. Here, \( C \) is the smoothing constraint operator chosen as the 2-D Laplacian [23]. This prefiltering operation tends to decorrelate the data in the spatial direction (by removing the blur), but leaves the relationships between wavelet coefficients in scale unchanged [40]–[42].

It is necessary to consider the effects of prefiltering on the noise in (32). In reality, the noise term \( \tilde{n}_F(s) \) is no longer white, but rather a colored noise whose spectral shape is dependent upon the blur and the choice of \( \alpha \). Since we are processing this new observation in multiscale space, however, the noise at each level of the wavelet decomposition may be considered locally white. That is, we can treat the new observed data at each subband \( i \) as the combination of some signal \( \tilde{f}_F(s) \) and a zero-mean white noise, \( \tilde{n}_F(s) \), with variance \( \sigma(i)^2 \tilde{n}_F \). The index \( i \) is an arbitrary index that we use throughout this paper to refer to specific subbands. For example, a three level 2-D decomposition would have nine wavelet coefficient subbands, and one lowpass subband, so \( i \) would range from one to ten.

We can approximate the local variances by incorporating our knowledge of the prefilter into the representation of this noise term. Writing the power spectrum of the prefiltered noise, \( S_{\tilde{n}_r}(\ell) \), in the spatial frequency domain, we can clearly see that at each 2-D frequency component \( \ell \),

\[
S_{\tilde{n}_r}(\ell) = \frac{|D(\ell)|^2 \sigma_n^2}{|D(\ell)|^2 + 2\alpha |D(\ell)|^2 |C(\ell)|^2 + \alpha^2 |C(\ell)|^2} \tag{36}
\]

where the original noise variance is \( \sigma_n^2 \). The shape of this function is generally bandpass in nature, with damped poles at the zero positions of the original blur function. A locally white assumption was also used for inverse filtered 1-D signals in the wavelet domain in [21]. In that work, an empirical measure was used at each level to obtain the noise variance. Here, we estimate it according to the average value of \( S_{\tilde{n}_r}(\ell) \) within the range of frequencies corresponding to each subband. For each subband \( i \), consider that the support in the frequency domain extends over \( L_i \) frequency components. The local noise variance of subband \( i \) is then computed according to

\[
\sigma(i)^2 \tilde{n}_F = \frac{1}{L_i} \sum_{j=1-i}^{L_i+i} S_{n_r}(\ell_j). \tag{37}
\]

We can use this as the estimate of the noise variance for each node on a tree of wavelet coefficients for the term \( \hat{h}_{in}(s) \) required by the multiscale filter in (19). However, it is clear that if the frequency band of interest includes components of the spectrum that lie near poles in this function, the estimate of the noise will be less accurate. The effect of these inaccuracies will be a biased estimate of the noise variance in that subband. The manifestation of this bias will then be an oversmoothing in the results, which will tend to bring the solution closer to the regularized smooth direct solution. In general, this estimate provides a reasonable range of values for the noise variance, though, and practical results support the use of this approach. Additional adaptivity to the characteristics of each tree of wavelet coefficients is provided through \( A(s) \) and \( B(s) \). The sensitivity to the values of these parameters is somewhat less than that of \( \sigma(i)^2 \tilde{n}_F \), because of the local variability of each. The parameter that has the largest effect on the multiscale Kalman filter gain, however, is \( B(s) \), since it represents the noise in the multiscale model, and, thus, the local confidence in the cross-scale predictions. Again, an overestimate of this model noise results in oversmoothing of the result. The estimation of these parameters is discussed in the next section.

B. Adaptivity Through a Multiple Model Approach

The choice of the AR parameter \( A(s) \) permits substantial adaptivity to the characteristics of the wavelet coefficients within a tree of coefficients. Because we are filtering the wavelet coefficients themselves, as opposed to the scaling coefficients, it is necessary to carefully examine the correlation between such coefficients in scale. There have been efforts to describe such correlations, although they have been primarily limited to characterizing the correlations between the wavelet coefficients of Brownian motion processes [17], [18]. First treating the processes associated with wavelet coefficients from different subband as stationary, it is possible to compute their correlation. The correlation between two coefficients along one orientation, \( w_{ij}(m, n) \) and \( w_{ik}(m', n') \), can be expressed generally [17] integrating over the support of the image as

\[
R_{w_{ij}}(j, k, m, n, m', n') = 2^{j+k} \int \int R_{ij}(x, x', y, y') \cdot \\
\psi^1(2^{j}x - m, 2^{j}y - n) \cdot \\
\psi^1(2^{k}x' - m', 2^{k}y' - n') \cdot \\
dxdydx'dy'. \tag{38}
\]

Thus, a stochastic model in scale relating these coefficients is dependent on the autocorrelation of the original image, \( f(x, y) \), and the wavelet basis functions used for each of the detail bands. Because \( \psi^1(2^{j}x - m, 2^{j}y - n) \) and \( \psi^1(2^{k}x' - m', 2^{k}y' - n') \) are orthogonal for \( j \neq k \), the correlation of wavelet coefficients from different scales will be largely negligible, or
at least decays rapidly depending on the number of vanishing moments of the wavelet basis [17]. The residual correlation that does exist lies in the transition bands of the wavelet bases. This correlation will reduce with higher order wavelets.

While this stationary model leads to a poor correlation between the coefficients of interest, a nonstationary approach is, in fact, a more appropriate model. The edge information may actually be highly correlated across scale (a characteristic that has received considerable attention in coding applications [33], [34], [43], and in denoising applications [7], [8]). To handle this problem, we utilize the multiple model approach. Because the separable 2-D wavelet transform provides coefficients representing different orientations of edges, we choose different AR parameters for quadtree from different orientations. In addition, because we assume that the edge regions of the image are highly correlated across scale, while the flat regions are not, we choose different AR parameters for “edge” and “non-edge” trees within each orientation. The choice of these parameters is based on an empirical measurement of correlations between sample sets of coefficients taken from different scales of trees from the different categories mentioned above.

In order to select the appropriate value of \( A(s) \) for the data in each quadtree, we first apply a detection algorithm that determines whether the data contained in each tree is dominated by noise or actual edge information. To do this, we model the distribution of the wavelet coefficients of the prefiltered image and the prefiltered noise in the coarsest scale, at one orientation, as zero-mean Gaussian [44], with respective pdf’s \( p^N_{x_p}(t) \) and \( p^E_{x_p}(t) \), with variance \( \sigma^2_{x_p}(t) \), for \( x_p \) and \( p^E_{x_p}(t) \) with variance \( \sigma^2_{x_p} \). We assume that the noise variance is known (as discussed in Section IV-A). The variance \( \sigma^2_{x_p} \) is also dependent upon scale (as designated by the subband index \( i \)) and must be estimated directly from the observed data, \( y \), at the coarsest scale. Because the passband of the blur contains the coarsest subbands in the cases considered here, and the SNR is high at this coarse scale, we assume that we can obtain a good estimate of \( \sigma^2_{x_p} \) based on \( \sigma^2_{x_p} \). Using a maximum likelihood detector, we take the two zero-mean distributions, and compute the optimal detection threshold. This is simply equal to the value of \( \tau \) where

\[
L(\tau) = \frac{p^E_{x_p}(t)}{p^N_{x_p}(t)} \bigg|_{t=\tau} = 1
\]

or

\[
\frac{1}{\sigma_x(t)_{f_p} \sqrt{2\pi}} \exp \left( \frac{-\tau^2}{2\sigma_x(t)_{f_p}^2} \right) = \frac{1}{\sigma_x(t)_{h_p} \sqrt{2\pi}} \exp \left( \frac{-\tau^2}{2\sigma_x(t)_{h_p}^2} \right)
\]

which yields

\[
\tau = \sqrt{\frac{\ln \frac{\sigma_x(t)_{h_p}}{\sigma_x(t)_{f_p}}}{\frac{\ln \frac{\sigma_x(t)_{h_p}}{\sigma_x(t)_{f_p}}}{2\sigma_x(t)_{f_p}^2} - \frac{1}{2\sigma_x(t)_{h_p}^2}}}.
\]

Using a different value of \( \tau \) for the three separate orientations at the coarsest scale as the detection threshold, a tree can be modeled as non-edge-dominated (\(|w^0_{x_p}(m, n)| < \tau \)), or edge-dominated (\(|w^0_{x_p}(m, n)| > \tau \)), for \( \lambda = 1, 2, \) or \( 3, \) depending on orientation, directly from the values of the observed wavelet coefficients at this scale (where the SNR is high).

The choice of the value of \( A(s) \) relies on making the following assumption. For a given class of quadtree, the nodes on these trees are each correlated with their parent nodes by the same amount. The classes are “edge” and “non-edge” within horizontal, vertical, and diagonal trees. Within each class of tree, a different \( A(s) \) may be measured for each scale change in the wavelet decomposition. Here we find the linear minimum mean square error estimate of \( A(s) \) for each scale change in each type of quadtree (edge or non-edge) using the noisy prefiltered data. First, within a particular quadtree, the following equation may be applied:

\[
\tilde{z}(s) = K \tilde{z}(s'\gamma).
\]

The best linear estimator of \( \tilde{z}(s) \) given \( \tilde{z}(s'\gamma) \) is obtained through

\[
\tilde{z}(s) = \frac{E\{\tilde{z}(s)\tilde{z}(s'\gamma)\}}{E\{\tilde{z}(s')\tilde{z}(s'\gamma)\}} \tilde{z}(s'\gamma).
\]

Using (32) and (33), this yields

\[
K = \frac{A(s)\sigma^2(t)_{f_p(s')}}{\sigma^2(t)_{f_p(s')} + \sigma^2(t)_{h_p(s')}\gamma}
\]

where \( \sigma^2(t)_{f_p(s')} \) indicates the variance of the edge or non-edge data in subband \( s' \) associated with the coefficient at position \( s \) in the quadtree, and \( \sigma^2(t)_{h_p(s')} \) indicates the variance of the prefiltered noise in subband \( s, \) which is associated with position \( s \) on the quadtree.

Consider now the data from one orientation only. Within this set of quadtrees, assume that the correlation between any edge (non-edge) coefficient at Scale \( j \) and any edge (non-edge) coefficient at Scale \( j-1 \) is the same. Then, coefficients from one class at a specific scale, \( j, \) are grouped into a signal called \( \tilde{z}_g(j) \), and coefficients from the previous scale, \( j-1, \) are grouped into a signal called \( \tilde{z}_g(j-1) \). Because the number of coefficients available at Scale \( j \) is four times as great as that at Scale \( j-1 \), it is only necessary to take one-fourth of the available samples in Scale \( j \) to create the signal \( \tilde{z}_g(j) \). This way, the two signals \( \tilde{z}_g(j) \) and \( \tilde{z}_g(j-1) \) will have the same length. The parameter \( K \) in (42) contains the parameter \( A(s) \), which we are trying to estimate. We can write the relationship governed by (42) in terms of this sample data as

\[
\tilde{z}_g(j) = K \tilde{z}_g^0(j-1) - 1
\]

Here \( \tilde{z}_g^0(j) \) is a single coefficient from the signal at Scale \( j \).

Finding

\[
A(s) = \arg\left\{ \min_{A(s)} \| \tilde{z}_g(j) - \tilde{z}_g(j) \|^2 \right\}
\]

yields

\[
A(s) = \frac{(\sigma^2(t)_{f_p(s')} + \sigma^2(t)_{h_p(s')}\gamma)\gamma}{(j - 1, j - 1)}
\]
where
\[ r(j, k) = \sum_n z_n^m(j) \tilde{z}_n^m(k). \] (48)

In (33), the appropriate value of the prediction parameter within a given class is set according to this estimate. This serves as an empirical estimate of the AR parameter for the cross-scale model at each node. The accuracy of this model will change with location, but it is possible to account for this changing accuracy by appropriately choosing the model error variance at each node.

Turning attention to \( B(s) \), from (32) and (33), we can write
\[ \tilde{z}(s) - A(s)\tilde{z}(s^+) = B(s)c(s) + \tilde{n}_F(s) - A(s)\tilde{n}_F(s^+). \] (49)

Given a good estimate of the parameter \( A(s) \) from the previous estimation technique, the model accuracy may be determined by choosing the \( B(s) \) which satisfies (49) in terms of the variance of both sides of this equation. Assuming zero-mean data in all cases, we have
\[ E\{(\tilde{z}(s) - A(s)\tilde{z}(s^+))^2\} = E\{(B(s)c(s) + \tilde{n}_F(s) - A(s)\tilde{n}_F(s^+))^2\}. \] (50)

This yields
\[ B(s) = \left( \sigma(i)^2 s(s) - 2A(s)\sigma(i)s(s)\tilde{z}(s^+) + \sigma(i)^2 s(s)A(s)\tilde{n}_F(s^+) \right) \right)^{1/2}. \] (51)

Again, the index \( i \) refers to the subband of interest at position \( s \) in the quadtree, however, we limit the data used in estimating the signal variances in (51) to be chosen exclusively from within each separate quadtree of wavelet coefficients. That way the model accuracy parameter \( B(s) \) is permitted to change with each quadtree in a very localized way. The influence of the model noise \( B(s) \) can be quite large. Because this method makes an estimate of the prediction parameter from noisy data, there can be some error in the model. This error term will cause amplification of the gain in the multiscale filter when it is large, thus permitting the influence of the innovation to be felt more strongly. This is very important for the edge sensitivity of the model. The model noise also tends to change with the different leaves on each quadtree, which is a result of the differing correlation between wavelet coefficients across scale.

C. Linking the Prefilter and Multiscale Filter

One of the principle issues in applying this restoration technique revolves around the choice of the prefiltering operation to perform the deconvolution prior to multiscale noise smoothing. From a heuristic viewpoint, it is clear that we would like to have the output of the prefilter provide as sharp an image as possible without too much amplified noise. This would suggest the choice of a relatively small regularization parameter, which would reduce the constraint on the image to be smoothed. On the other hand, the framework in which this problem is cast provides the means to analytically choose the output from multiple prefilters whose restorations are based on different choices for the regularization parameter \( \alpha \) in (35).

Recalling that it is necessary to perform a detection operation in the coarsest subband along each orientation, we must require that the signal variance be larger than the noise variance here, after prefiltering, in order to allow for the accurate classification of quadrants into edge and non-edge groups. To do this, we must find the appropriate value of \( \alpha \), for which each coarse subband \( i \) results in
\[ \sigma(i)^2 \leq \sigma(i)^2 \] (52)

Using (37), this becomes a straightforward operation, given an estimate of \( \sigma(i)^2 \) based on the variance of the observed subband \( \sigma(i)^2 \). This estimate is valid because the passband of the blur contains the coarsest subbands, and the SNR is high here as well. Evaluating the amplification of the original noise variance in a particular subband as a function of \( \alpha \), and setting a threshold based on the estimated signal variance at a given blurred SNR according to (52), we can choose the lower bound for each subband. From (37), this bound satisfies
\[ \frac{1}{L_i} \sum_{j=1+i}^{L_i+i} \left| D(j) \right|^2 \sigma_n^2 + 2\alpha|D(j)|^2 |C(j)|^2 + \alpha^2 |C(j)|^4 = \sigma(i)^2 \] (53)

or
\[ \frac{1}{L_i} \sum_{j=1+i}^{L_i+i} \left| D(j) \right|^2 |D(j)|^4 + 2\alpha|D(j)|^2 |C(j)|^2 + \alpha^2 |C(j)|^4 \right) + \sigma(i)^2 \] (54)

Fig. 5 shows an example of the intersection of three constants (the second term on the left hand side of (54)) with a noise curve (the first term on the left hand side of (54)) for the coarsest subband of the (a) HH, (b) LH, and (c) HL orientations. These curves apply to the restoration of the 256 \( \times 256 \) Cameraman image (seen in Fig. 4 (a)), blurred by a uniform 9 \( \times 9 \) blur at various noise levels. The intersection of each straight line with the curved line gives a lower-bound choice for the value of \( \alpha \) at that BSNR, which will permit accurate edge and non-edge tree detection in the multiscale filter. The line at 40 dB BSNR does not intersect with the noise curve in Fig. 5 (b) and (c), indicating that a lower-bound choice for \( \alpha \) in these instances would be zero. The lower bound values of \( \alpha \) chosen for each of the coarse subbands in this way may be used to prefilter the observed image \( y \) and perform the edge and non-edge tree detection in the subbands of each of these prefiltered images. The result of this detection is then used as an edge map for the subsequent multiscale filtering operation. In other words, we have considered applying a different prefilter to the original image multiple times. Each time, the choice of \( \alpha \) is based upon a particular subband, and that subband from the wavelet-decomposed resulting image is used to compute an edge-map.

Sometimes, choosing \( \alpha \) based on the coarse subbands leads to such small \( \alpha \) values that the finer scale subbands become
buried in noise. This makes the application of the multiscale noise smoothing filter impractical. In the same way, choosing $\alpha$ based on the coarse subbands may lead to some subbands in the finest scale becoming over-smoothed, which will reduce the edge sensitivity desired in the result. As a means of protecting against these problems, after detection has been performed, a second set of prefiltered images may be used as the actual data to be processed by the multiscale filter. The values of $\alpha$ chosen for the prefilters used to generate these images are based on the estimated noise and signal variances in the finest subbands. Fig. 6 shows the same type of plot as that in Fig. 5 for the finest (a) HH, (b) LH, and (c) HL subbands, in a three-level decomposition. Here again the amplification to the original noise variance after prefiltering is given by the curved line as a function of $\alpha$. The straight intersecting lines represent optimal values based on an estimated signal-to-noise constant in each subband at different BSNR's. This optimal constant is chosen to be $\frac{1}{10}$ of that chosen for the corresponding coarsest scale subband plot. This is based on the knowledge that the power of white noise in the originally observed subbands increases by four times for every jump in scale from coarse to fine. In addition, we have a good estimate of the signal variance in the coarse subbands, but not in the finest subbands. This is because the observed subbands at the finest scale are not contained in the passband of the blur. Thus, information from the coarse scale is used as a basis for picking an $\alpha$ for the fine scales as well.

V. RESULTS

As global measures of objective improvement for the restoration techniques discussed here, we refer to both the improvement in SNR (ISNR), given by

$$ISNR = 10 \cdot \log_{10} \left( \frac{\sum_{m,n} [f(m,n) - y(m,n)]^2}{\sum_{m,n} [f(m,n) - \hat{f}(m,n)]^2} \right)$$

(55)

and the mean square error (MSE) given by

$$MSE = \frac{1}{N^2} \sum_{m,n} [f(m,n) - \hat{f}(m,n)]^2$$

(56)

for an $N \times N$ image, where $f(m,n)$ and $y(m,n)$ are the original and degraded intensity components, respectively, at spatial location $m, n$, and $\hat{f}(m,n)$ is the corresponding restored intensity field. Although these measures reflect the global properties of the restoration and may not fully reflect the subjective improvement of spatially adaptive approaches, they are nonetheless useful to include it here to provide an objective means to measure and compare the quality of the results.
As an initial presentation of the performance of this algorithm, it is instructive to look at the restoration of an image in the presence of noise only. This situation may be encountered in a number of imaging scenarios where there is little or no degradation due to defocusing or motion. Given an image that is degraded by noise only, we can apply the algorithm described in this paper directly, simply by treating the prefiltering operation as an all-pass (or identity) operator, such that the noise removal of the multiscale Kalman filter is demonstrated independently of the deblurring operator. This is accomplished by setting $F = I$ in (35), and proceeding with the multiscale filtering. As an illustration of this concept, Fig. 4(b) shows the Cameraman image degraded by additive white Gaussian noise at an SNR of 10 dB. Applying the wavelet-domain multiscale smoothing filter described in Section III to this image, Fig. 4(c) shows the filtered version of this image, having an ISNR=3.41 dB. For this high additive noise case, the signal variance estimates are difficult to obtain from the observed coarsest subbands, thus we enforce the rule that the empirically estimated signal variances for the noisy trees be less than those of for the edge trees here. It is clear that the spatially adaptive filtering used here preserves the edges in the image, while reducing the noise dramatically in the homogeneous regions of the image.

When considering the case in which blur is present, we look at a number of comparisons to understand the benefits of the proposed method. First, in order to compare and contrast the performance of the proposed algorithm with some of the standard direct approaches discussed in Section II, we show the restoration of the blurred Cameraman image with both a Wiener filter and a CLS filter. The parameters used in each of these filters are chosen to demonstrate a worst-case
scenario, when we have very little prior information about the original image. For the Wiener filter, we use an estimate of the required power spectrum obtained from the observed degraded image. For the CLS direct filter, we use a regularization parameter $\alpha$ equal to the reciprocal of the blurred SNR of the observed image. The choice of $\alpha = \frac{1}{BSNR}$ can be obtained through the set-theoretic formulation of the problem, where the loose bounds on the constraints we try to minimize are the noise variance and the signal variance [25]. Here, the SNR is assumed to be known, or can be estimated from the available data [24]. The purpose of choosing these parameters for the direct filters is to demonstrate the shortcomings of the nonadaptive approaches in this situation, and to show how the wavelet-based approach is able to outperform these direct methods quite well, given the same prior knowledge.

Applying separate prefilters to the degraded image as described in Section IV-C and then using the subbands of those prefiltered images, a multiscale adaptively restored image was generated for each of the cases tested here. Here Table I shows the optimal lower bound choices for the values of $\alpha$ in the prefilter as determined by each of the coarsest and finest subbands for the 256 × 256 cameraman image degraded by a uniform 9 × 9 blur. Fig. 7 shows the results for this image at 40 dB BSNR. Fig. 7(a) is the degraded image, Fig. 7(b) is the Wiener restoration, Fig. 7(c) is the CLS restoration, and Fig. 7(d) is the wavelet-based adaptive restoration. The wavelet domain data was acquired by passing the image through a three-level separable decomposition using the Daubechies 16-tap wavelet [45]. We compensate for the asymmetry of 16-tap wavelet by circularly shifting each subband to align correlated edges properly. This wavelet is chosen because of its high regularity properties. The smoothness provided by this decomposition results in fewer highly structured components in the finer subbands, and thus better performance when filtering out the
residual noise present there. Shorter wavelets tend to give somewhat noisier results for this image restoration application.

As a means of comparing this technique to existing spatially adaptive techniques in the literature, we test the performance of this algorithm against a spatially adaptive iterative least-squares algorithm constrained by a visibility function. This type of approach was introduced in [28], which formed the basis for a number of nonstationary approaches thereafter. In [27], the author uses the method of generalized cross-validation (GCV) to obtain a spatially adaptive image restoration based on the iterative constrained least-squares algorithm. In that paper, results were obtained for the Building image using no a priori information about the original image. Here, those results are compared to the results obtained for the same degraded images with the multiscale restoration procedure presented in this paper. The blur operator used here is an out-of-focus blur modeled by

\[
d(i, j) = \begin{cases} \frac{1}{\tau^2}, & \text{if } \sqrt{i^2 + j^2} \leq R \\ 0, & \text{otherwise} \end{cases}
\]

with \( R^2 = 17 \). The results in [27] were presented in terms of MSE, and Table II shows the resulting MSE’s for both the GCV and the proposed multiscale adaptive method for 20-, 30- and 40-dB BSNR. The proposed approach has an MSE less than that of GCV for the 30-dB case, and slightly more for the other two cases. It is important to note the difficulty in using global metrics to measure improvement when spatially adaptive techniques are used. Nonetheless, it can bee seen that the noniterative multiscale method is quite comparable to the iterative method presented in [27]. The original Building image, the degraded image at 30-dB BSNR, and the restored version using the multiscale technique proposed here are shown in Fig. 8(a)–(c).
Given that this example demonstrates the comparability of the proposed approach with iterative procedures, it is important to point out the computational advantages of the noniterative wavelet-based approach. From a computational complexity viewpoint, the proposed algorithm requires only limited expense over the direct methods discussed above, and is in general less computationally expensive than the iterative approach, or comparable iterative procedures, described in this section.

The prefiltering procedure that is applied for the deconvolution operations may be simply applied in the discrete frequency domain. One prefiltered image is generated for each of the regularization parameters for the three orientations. The appropriate subbands from each prefiltered image are then extracted through the wavelet-based filtering of these reconstructed images. A three-level dyadic wavelet decomposition is used, so there are ten total subbands to be generated. The operations then performed by the multiscale filter are quite fast, especially when considering the independent nature of each quadtree of wavelet coefficients. In this case, we have quadtrees made up of 21 coefficients each, which may all be processed in parallel. The cost of the multiscale filter in relation to the forward and inverse transforms is negligible. The multiple prefiltering operations required by this procedure (six were used here) thus comprises the bulk of the computational expense.

For iterative restoration procedures such as those described in [27] and [46], the bulk of the computational complexity lies in the multiple iterations of the basic steps in successive approximation. In order to incorporate spatially varying prior information about the solution, iterative techniques solve the deconvolution (inverse) problem by successively updating an approximation of the solution, using some type of projection of the previous solution onto a constraint space. Direct inversion is not useful for such an approach. The cost of performing these operations depends on the projection technique being used, but usually solutions requiring several hundred iterations will take significantly longer than the proposed method, which uses only a small number of direct transforms coupled with parallel recursive filtering in the wavelet domain.

<table>
<thead>
<tr>
<th>BSNR (dB)</th>
<th>MSE (GCV)</th>
<th>MSE (Multiscale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>270.6</td>
<td>293.9</td>
</tr>
<tr>
<td>30</td>
<td>180.8</td>
<td>176.6</td>
</tr>
<tr>
<td>40</td>
<td>76.1</td>
<td>78.60</td>
</tr>
</tbody>
</table>
From a qualitative viewpoint, it is apparent that the multiple prefilters used by the multiscale method provide the means to accurately determine the presence of edge and non-edge trees, without allowing too much noise to pass through the prefiler. The results found here and in other experiments show that this technique provides very good preservation of edges even for such severe degradations as those shown here. These results also demonstrate the effectiveness of this adaptive approach in comparison to the direct filters, and to spatially adaptive approaches as well.

VI. CONCLUSIONS

Multiscale wavelet domain approaches to image restoration provide a number of distinct advantages over spatial domain approaches. The most important of these is that the edges in an image may be accurately preserved. Because of the perceptual importance of these edges, any multiscale adaptive technique that does not increase complexity too much over standard approaches can be very beneficial.

In this paper, we have proposed a new approach to spatially adaptive image restoration, which has a minimal additional computational load in comparison to direct techniques. The use of the wavelet coefficients here presents a new method for adaptive restoration, and leads to very good edge preservation in the restored results. This is a benefit of the fact that the wavelet coefficients may be interpreted as representations of the edges in an image at different scales and orientations. The formulation of the problem in the wavelet domain leads to an understanding of the relationships between coefficients across scales and the behavior of the linear blur operator in this domain. Because the multiscale state space representation in this domain is difficult to work with, we found that prefiltering the data with a decorrelating filter led to an improved multiscale part of the restoration allows us to adapt the solution to the edges in the wavelet domain, and to prevent excessive noise amplification in the result.

The multiscale filter implemented here provides the majority of noise smoothing required. It should also be noted that the formulation provided here supports the use of prefilters different from the simple direct CLS filters used here. Any filtering technique that provides a sharp deconvolution and decorrelates the multiscale state vectors may be used. The technique presented here was chosen because of its optimality for the required detection problem in the multiscale filtering step. In addition, the multiscale implementation of the noise-smoothing filter is very computationally efficient, adding little complexity over the direct deconvolution prefilters in order to gain spatial adaptivity.

ACKNOWLEDGMENT

The authors wish to acknowledge the many valuable comments provided by the anonymous reviewers of this paper.

REFERENCES


Mark R. Banham (S'87–M'94) was born in Philadelphia, PA, in 1966. He received the B.S.E.E. degree from Drexel University, Philadelphia, in 1989, and the M.S. and Ph.D. degrees in electrical engineering from Northwestern University, Evanston, IL, in 1990 and 1994, respectively.

While at Northwestern, he was a Murphy Fellow, a teaching assistant, a research assistant, and an instructor in the Electrical Engineering and Computer Science Department. Since 1994, he has been a senior engineer with the Digital Technology Research Laboratory at Motorola's Corporate Research Laboratories, Chicago, IL, USA, working on low bit rate video compression for personal video communications. His research interests are in the general areas of image coding and restoration.

Dr. Banham is a member of Tau Beta Pi andEta Kappa.

Akgelos K. Katsaggelos (S'80–M'85–SM'92) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees both in electrical engineering from Georgia Institute of Technology, Atlanta, in 1981 and 1985, respectively.

In 1985 he joined the Department of Electrical Engineering and Computer Science at Northwestern University, Evanston, IL, where he is currently an associate professor. During the 1986–1987 academic year he was an assistant professor at Polytechnic University, Department of Electrical Engineering and Computer Science, Brooklyn, NY. His current research interests include image recovery, processing of moving images (motion estimation, enhancement, very low bit rate compression), and computational vision.

Dr. Katsaggelos is an Ameritech Fellow and a member of the Associate Staff, Department of Medicine, at Evanston Hospital. He is a member of SPIE, the Steering Committees of IEEE TRANSACTIONS ON MEDICAL IMAGING and IEEE TRANSACTIONS ON IMAGE PROCESSING; the IEEE Technical Committees on Visual Signal Processing and Communications and on Image and Multidimensional Signal Processing; the Technical Chamber of Commerce of Greece; and Sigma Xi. He has served as an associate editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING (1990–1992), and an area editor for Graphical Model's and Image Processing (1992–1995). He is the Editor of Digital Image Restoration (Springer-Verlag, Heidelberg, 1991) and was the General Chairman of the 1994 Visual Communications and Image Processing Conference, Chicago, IL.