

Image Restoration Using Space-Variant Gaussian Scale Mixtures in Overcomplete Pyramids

Jose A. Guerrero-Colón^{*†}, Luis Mancera[†] and Javier Portilla[♣]

Abstract—In recent years Bayes Least Squares - Gaussian scale mixtures (BLS-GSM) has emerged as one of the most powerful methods for image restoration. Its strength relies on providing a simple and yet very effective local statistical description of oriented pyramid coefficient neighborhoods via a GSM vector. This can be viewed as a *fine* adaptation of the model to the signal variance at each scale, orientation and spatial location. Here we present an enhancement of the model by introducing a *coarser* adaptation level, where a larger neighborhood is used to estimate the local signal covariance within every subband. We formulate our model as a Bayes least squares estimator using space-variant Gaussian scale mixtures. The model can be also applied to image deconvolution, by first performing a global blur compensation, and then doing local adaptive denoising. We demonstrate through simulations that the proposed method, besides being model-based and non-iterative, it is also robust and efficient. Its performance, measured visually and in L2-norm terms, is significantly higher than the original BLS-GSM method, both for denoising and deconvolution.

Index Terms—Image restoration, image denoising, overcomplete oriented pyramids, Gaussian scale mixtures, Bayesian estimation

I. INTRODUCTION

IMAGE restoration is typically formulated as the estimation of an image given a linearly filtered version of the original corrupted by additive noise. This is a difficult, ill-posed, inverse problem, even if, as assumed in this work, the degradation parameters are known. A more tractable problem arises when we only consider additive noise in the degradation model (denoising problem). During the last few years, the state of the art in image denoising has experienced a formidable advance. The possibility of using image denoising techniques within more generic image restoration frames is thus becoming more and more attractive. In this paper we present a new denoising-based approximation to deal with both degradation problems (just noise, and blur plus noise). For previous instantiations of this work see [1], [2].

Wiener filtering is one of the simplest model-based restoration methods, and yet it provides optimal results when dealing

with Gaussian signals and noise. Modelling noise as Gaussian, as we do in this work, is reasonable for many image processing situations. However, it is well known that image statistics are far from Gaussian [3]. One can, instead, assume that images are *locally* Gaussian. Spatially adaptive models have widely used this assumption for image denoising. In 1980, Lee [4] proposed an adaptive scalar Wiener method in the pixel domain, based on estimating the local variance at every spatial location of the image. With the introduction of *wavelets* (in the broad sense of the word, i.e., multi-scale, multi-orientation linear pyramidal representations), local adaptation in the new domain became much more powerful. Whereas many authors explored other estimations approaches in the new domain, which resulted in different types of shrinkage functions to be applied to the wavelet coefficients (e.g. [5], [6], [7], [8]), others continued developing the local Wiener idea. Mihçak et al. [9] included a prior on the local variance, proposing a maximum a posteriori (MAP) estimation scheme. A similar Bayesian approach was followed in [10]. Portilla et al. [11] refined the Bayesian model by including the full signal and noise covariance in the model, and simplified the formulation of the problem by using a Gaussian scale mixture (GSM) [12], [13] frame [14]. These Wiener adaptive methods followed an empirical Bayes strategy: first, estimate the local signal variance, and then, apply a Wiener filter using that estimation. In a further work, Portilla et al. [15], instead, introduced a single-step Bayes Least Squares (BLS) estimator, which resulted in a significant performance increase. Up to now, all Wiener-adaptive methods either consider just the local variance, disregarding the covariance structure for each subband [9], [10], or consider the covariance fixed, up to a scale factor [11], [15]. In both cases they perform a *fine* adaptation, in the sense that they estimate the local variance using small neighborhoods. On the other hand, there is broad literature on local covariance estimation and its applications (e.g. [16]). Covariance estimation allows for a selective characterization of salient local spectral features. However, as it requires a much larger window than variance estimation, it loses the ability to represent rapid variance spatial fluctuations in small areas. As a result, for typical images, the extra complexity of a covariance-adaptive model, by itself, does not pay off in performance terms with respect to simpler variance-adaptive models. Here we describe a new approximation that provides both kinds of adaptation in the same model [1]. Hammond and Simoncelli [17] have also proposed an interesting extra adaptation step for the BLS-GSM model, in their case constrained to the local dominant orientation.

[†] Department of Computer Science and Artificial Intelligence, Universidad de Granada, Spain. Email: {jaguerrero, mancera}@decsai.ugr.es.

[♣] Department of Images and Vision, Instituto de Óptica, CSIC, Madrid, Spain. Email: portilla@io.cfmac.csic.es.

JAGC is supported by AP2005-2969 FPU contract and TEC2006/13845/TCM grant, both funded by the Ministerio de Educación y Ciencia (Spain).

LM is supported by TIC2003-01504 FPI contract and TEC2006/13845/TCM grant, both funded by the Ministerio de Educación y Ciencia (Spain).

JP is supported by TEC2006/13845/TCM grant and the Ramón y Cajal Program, both funded by the Ministerio de Educación y Ciencia (Spain).

BLS-GSM [15] has provided the state-of-the-art in image Gaussian denoising for the last few years. During the last year, however, some new methods have achieved impressive results, both visually and in L2 norm terms, some of them clearly surpassing the BLS-GSM performance. Many exploit an important characteristic of typical image: they present a high spatial redundancy in the sense that a local feature is likely to be found in several (possibly disconnected) spatial locations within an image (e.g., a corner of a window, in a building façade, a straw bar, or, more simply, part of an edge along the edge line). An important part of the differences between these repeated local features may be caused by noise, so this fact can be exploited for noise removal. The usual approach is to consider that every neighborhood of pixels can be written as a linear combination of other neighborhoods. Buades et al. [18] do this matching for every pixel location, using the contextual similarities among neighborhoods around the pixels. Kervrann and Boulanger [19] improve this method by introducing an adaptive neighborhood as well as a statistical estimation of the model parameters. Hirakawa and Parks [20] use a total least squares formulation for square neighborhoods of different sizes. Dabov et al. [21] use block-matching and filter the matched blocks in 3D. Once these blocks are 3D-filtered, they are denoised by a combination of hard-thresholding and local Wiener filtering. To the best of our knowledge these authors achieve the current state of the art in performance terms. However their method is not based on a single model for the estimation, that is, it applies re-estimations. We refer here to *re-estimation* as when the result of an estimation method is used as the input data for a *different* method with the *same* estimation purpose¹. Though possibly effective and convenient in some practical situations, re-estimation is not a conceptually clean practice, as it implies the overlapping of two mutually incoherent models. Another successful path to image denoising is to adapt to the geometry of the image. Foi et al. [22] have presented a method based on a shape-adaptive discrete cosine transform (SA-DCT), which adapts the local support for the data being processed to the image contents. They use the same re-estimation method as in [21]. Elad et al. [23] achieve impressive results as well by creating an image-adaptive sparse representation. They used the K-SVD algorithm to obtain a dictionary that describes the image content very effectively. A big part of the high performance of this latter method is due to using a non-linearly adaptive image representation, which requires a heavy computational load.

In contrast to the methods referred above, the one we propose here can use any linear multi-resolution representation (though best results are obtained with multi-oriented translation invariant pyramids) and it does not make use of re-estimation. The underlying model is very simple, and the resulting method is non-iterative, efficient and robust, inheriting and extending on the good properties of the original BLS-GSM, such as its ability to deal with any spectral noise density.

When approaching the deconvolution problem, we can see

¹When both estimation methods are the same, we talk about iterative estimation, which is a conceptually clean practice (e.g., the E-M approach).

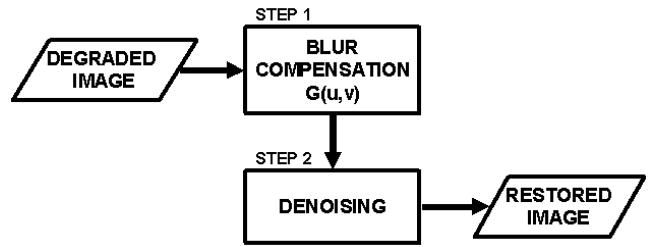


Fig. 1: Two step scheme.

that most methods are iterative and global, e.g. [24], [25], [26], [27]. Among the global methods, some authors (e.g. [28], [29], [30], [31], [32], [33]) have approached the deblurring problem by decomposing it in two steps (see Fig. 1), *a*) a linear global compensation of the blur in the frequency domain, and *b*) a non-linear local denoising method. We emphasize that this procedure is not a form of re-estimation, because the problem addressed in the first step is different from the one addressed in the second step. The advantage of this decomposition is that it converts the deconvolution problem into an easier denoising problem, opening the door to recent powerful non-iterative methods whenever they are able to represent and remove noise with arbitrary spectral density (PSD). Note that, by default, classical denoising assumes white noise. In contrast, here, even if the original noise is white, it gets colored by the blur compensation filtering in the first step.

Few authors instead, e.g. [34], [35], have approached the deconvolution problem using local models, both for the image and for the degradation, with the advantage of simplifying the formulation of the problem. Portilla and Simoncelli [35] used a local GSM model within an efficient non-iterative method which provided good performance in most cases. However, the local blur compensation used in this method has some intrinsic limitations (see Section III-B). Also, its performance depends critically on an accurate estimation of the power spectrum of the original image.

In this work we propose a non-iterative two-step deblurring method based on the previously described spatially adaptive denoising. It provides a sub-optimal, but very robust and efficient, choice for the pre-filtering kernel. This approach preserves the advantages of using a local model for image denoising (simple and non-iterative) whereas it overcomes the problems associated to locally compensate for the effects of convolution. It also increases both the performance and the robustness of its predecessors BLS-GSM restoration methods.

The outline of this paper is as follows: Section II describes our image statistical modelling, including a description of the image representations used. Section III explains how to use our model to do both denoising and deblurring. Implementation details are reported in Section IV. Section V presents and discusses a broad set of simulations and their corresponding results. Last section concludes the paper.

II. IMAGE MODELLING

A. Image representation

Image representation is a key issue for image processing, and its choice depends on the particular application. For

image restoration it has been widely demonstrated (see, e.g. [36], [37], [15]) that translation invariant, over-complete representations provide far better results than non-redundant representations, such as critically-sampled wavelets. Among them, those based on a multi-scale analysis of the image (usually in a pyramid format), with orientation-selectivity typically provide the best results. Examples are the steerable pyramid [38], the curvelets [39] or the complex wavelet [40], among others. However, it has also been observed that each representation scheme fits best for processing certain image features, but none seems superior in absolute terms (see, e.g. [41]). Some authors have proposed to combine them by averaging their results [42], [1], although more powerful techniques are based on a non-linear basis selection using a sparseness criterion [41]. A different approach is to use image content adaptive basis functions such as wedgelets [43], bandlets [44] or K-SVD overcomplete dictionaries [23], among others. In this work, we have used two representations exhibiting complementary features: the Full Steerable Pyramid (FSP) [15] and a translation invariant version of the Haar wavelet [1] that we have termed Translation Invariant Haar Pyramid, TIHP.

1) *The Full Steerable Pyramid*: The FSP [15] is an extension of Simoncelli and Freeman’s original steerable pyramid [38] that splits the original high-pass residual into the same number of orientations as the rest of the pyramid levels. Some good properties of the FSP are that *a*) their filters can be made highly selective in frequency and *b*) it is flexible in its number of orientations. Its drawbacks are coupled with these advantages: *a*) due to their high spectral selectivity, its filters have a large spatial support and *b*) due to the possibly very high number of subbands, it has a relatively high computational cost (best denoising results with BLS-GSM are reported for 8 orientations [15], which results in a redundancy factor close to 19). According to these features, FSP is more suitable for processing texture-rich images.

2) *Translation Invariant Haar Pyramid*: The TIHP is an overcomplete version of the Haar wavelet [45]. It is translation invariant and it provides almost perfect reconstruction. For a detailed description see Appendix A. In contrast to FSP, TIHP’s kernels are very small, providing excellent spatial localization. Its moderate redundancy factor (7) and reduced computational cost, make it fast and easy to operate with. TIHP’s main weaknesses are its poor oblique orientations selectivity and the low spectral localization of its kernels. These features made this representation suitable for images with little texture, most of the energy being concentrated along edges and lines (especially if oriented horizontally or vertically).

B. Image statistics and Gaussian scale mixtures

Image statistics were first described in terms of its spectral properties. Empirical studies showed that typical behavior followed a $f^{\nu-2}$ law, being ν a small correction factor [46],

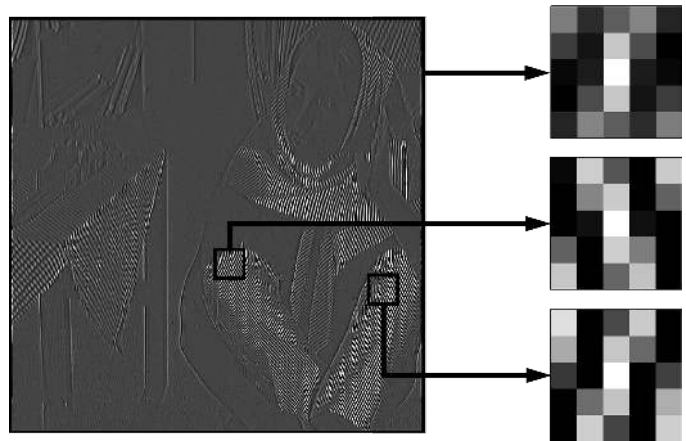


Fig. 2. Autocovariance variability within Barbara TIHP vertical band. We assume a 3×3 neighborhood and compute each autocovariance patch using samples from the associated bounded region.

[47]. Thus, usual approximations, fifty years ago, were to consider images as samples from highly correlated Gaussian random fields. However, in 1989 Mallat [3] showed with empirical data that band-pass responses to typical images are far from Gaussian: its marginal probability density functions have a pronounced peak at zero, and heavy tails. This caused the sparse behavior of the wavelet coefficients: only a small proportion of them concentrates a high proportion of the total energy, and this occurs not only for the whole representation, but also for every subband. The histograms reflect high amplitude responses caused by most relevant localized image features (such as edges, lines, corners, etc.), whereas spatially dominant responses are typically small in amplitude, mainly corresponding to low-contrast texture within homogeneous patches in the scene. Therefore, the dominant source for non-Gaussianity of image statistics is the inhomogeneity of the image contents, which causes abrupt changes and localized high amplitude responses at the objects’ boundaries. In this way, it was definitely assessed the inadequacy of using global Gaussian distributions to model typical images. In 1997, Simoncelli [48] showed another critical observation about image statistics: there is a strong coupling among neighbor *wavelet* coefficients in amplitude. This revealed explicitly a fundamental non-Gaussian joint statistical behavior that was coherent with the marginal densities studied so far. This feature has been exploited for image compression [49], [50], texture modelling and synthesis [51], and image restoration (all the predecessors of this work), among other applications. Two years later, Wainwright and Simoncelli found that Gaussian scale mixtures, GSM [12], was the proper mathematical tool to capture both the sparseness and the amplitude coupling of wavelet coefficients [14]. GSM’s have typically been used as a local description for neighbor coefficients in a pyramid (e.g. [15]), but they can also be used in a more powerful (but also more complex) global statistical description of the image in the transformed domain [52].

The GSM model in the wavelet domain [14] is able to represent the observed strong local variance fluctuations in neighborhoods of coefficients at the different spatial locations, scales and orientations. This spatial inhomogeneity of the

local variance explains, on its own, the two basic non-linear statistical patterns mentioned above, namely, sparseness of the responses and amplitude coupling among neighbors. The original GSM model assumed a fixed normalized covariance for all the neighborhoods at a given subband. However, it is clear that typical images have different spatial correlation patterns at different locations in a subband, even if the filtering associated with each subband greatly reduces the spatial variability of the local covariance in the image domain. We illustrate this in Fig. 2, where, using a TIHP vertical subband of *Barbara*, we have compared the local auto-covariance samples estimated for the whole subband (assuming a 3×3 neighborhood) with that computed from two different regions, both having high-contrast features. In this case the corresponding auto-covariance samples reveal significant differences in local orientation. Note that the auto-covariance of the whole subband reflects the average behavior, reducing the potential of a more spatially selective signal/noise discrimination. Thus, to further increase the realism and performance of our model, it is justified to use a Gaussian scale mixture with a spatially adaptive covariance for the local signal, that is, a space-variant GSM.

Let's consider an image decomposed into oriented subbands at multiple scales. We denote as $x_c^j(n, m)$ the coefficient corresponding to the subband j , centered at spatial location (n, m) . We denote as $\mathbf{x}^j(n, m)$ a *neighborhood* of coefficients clustered around this *reference coefficient* (see neighborhood structure in section IV). We model these neighborhoods as Gaussian scale mixtures (GSMs). A random vector \mathbf{x} is a GSM [12] if it can be expressed as the product of two independent random variables:

$$\mathbf{x} = \sqrt{z}\mathbf{u},$$

where z is a positive scalar and \mathbf{u} is a zero-mean Gaussian vector. We model the desired space variant behavior by defining $\mathbf{C}_{\mathbf{u}}^{j,(n,m)}$ as the (space variant) covariance of $\mathbf{u}^j(n, m)$. The density of \mathbf{x} is determined uniquely by $p_z(z)$ and $\mathbf{C}_{\mathbf{u}}^{j,(n,m)}$ as follows:

$$\begin{aligned} p^{j,(n,m)}(\mathbf{x}) &= \int p^{j,(n,m)}(\mathbf{x}|z) p_z(z) dz \\ &= \int \frac{\exp\left(-\mathbf{x}^T(z\mathbf{C}_{\mathbf{u}}^{j,(n,m)})^{-1}\mathbf{x}/2\right)}{(2\pi)^{N/2}|z\mathbf{C}_{\mathbf{u}}^{j,(n,m)}|^{1/2}} p_z(z) dz, \end{aligned}$$

where N is the dimensionality of \mathbf{x} and \mathbf{u} (the size of the neighborhood). Note that $\mathbf{x}^j(n, m)$ is conditionally Gaussian for a given z . There are different possibilities for the mixing density $p_z(z)$, which are described in [15]. As in the latter work we have assumed it constant in $\log z$ for all the subbands in the working range, which corresponds to a non-informative Jeffrey's prior.

III. IMAGE RESTORATION

A. Denoising using space-variant GSMs

We aim to solve the classical denoising problem, where an image is corrupted by additive zero-mean independent Gaussian noise of known (but arbitrary) spectral density. As

many other methods, we perform the denoising in the pyramid domain and obtain the image estimation by reconstructing the image from the estimated coefficients. The observation model is, in this case:

$$\mathbf{y} = \mathbf{x} + \mathbf{w},$$

where \mathbf{y} , \mathbf{x} and \mathbf{w} are vectors (in a lexicographic order) of observation, original and noisy wavelet coefficients, respectively.

1) *Two-level adaptive denoising*: Note that the model in [15] used a fixed signal covariance at each subband, providing only (fine) adaptation to the local variance (through the GSM hidden multiplier). Here we include another level of adaptation (coarse) by also letting the signal covariance to be a function of the spatial position at each subband. Following the GSM model presented in section II-B, we can express a neighborhood of N noisy coefficients as²:

$$\mathbf{y} = \sqrt{z}\mathbf{u} + \mathbf{w},$$

where \mathbf{w} is the zero mean Gaussian noise vector with covariance matrix $\mathbf{C}_{\mathbf{w}}$. Note that $\mathbf{C}_{\mathbf{w}}$ (unlike the signal covariance $\mathbf{C}_{\mathbf{u}}$) is assumed constant for all neighborhoods within a subband (but different for each subband). By GSM definition z and \mathbf{u} are independent, and by assumption \mathbf{w} is independent too. Hence the density of the observed neighborhood vector conditioned on z is Gaussian (zero-mean), with covariance $\mathbf{C}_{\mathbf{y}|z} = z\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}}$:

$$p(\mathbf{y}|z) = \frac{\exp(-\mathbf{y}^T(z\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}})^{-1}\mathbf{y}/2)}{((2\pi)^N|z\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}}|)^{1/2}}. \quad (1)$$

2) *Parameter estimation*: Given equation 1, signal and noise covariances are needed. The neighborhood noise covariance, $\mathbf{C}_{\mathbf{w}}$, is computed at each subband by applying the pyramidal representation to a deterministic function with the same sample autocorrelation as the noise (for instance, a scaled delta function for white noise). The elements of $\mathbf{C}_{\mathbf{w}}$ are computed by averaging the products of pairs of coefficients over all the neighborhoods of the subband. Note that, unlike noise, this deterministic function yields a result that is free from random fluctuations.

The neighborhood observation covariance, $\mathbf{C}_{\mathbf{y}}$, can be computed from $\mathbf{C}_{\mathbf{y}|z}$ by taking the expectation over z :

$$\mathbf{C}_{\mathbf{y}} = \mathbb{E}\{\mathbf{C}_{\mathbf{y}|z}\} = \mathbb{E}\{z\}\mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}}.$$

We can set, without loss of generality, $\mathbb{E}\{z\} = 1$ obtaining $\mathbf{C}_{\mathbf{y}} = \mathbf{C}_{\mathbf{u}} + \mathbf{C}_{\mathbf{w}}$. To estimate the spatially variant signal covariance $\mathbf{C}_{\mathbf{u}}$ we assume that it changes smoothly over the image, and, thus, that it may be sampled by estimating it into a set of non-overlapping regions³. We divide each subband into $B \times B$ blocks, which are assumed uniform in covariance.

²For notational simplicity from here on we drop the superscript j and the indices (n, m) . Note that the following equations are the same as in [15], but now $\mathbf{C}_{\mathbf{u}}$ represents a local covariance, as it depends on the spatial position.

³We have empirically tested that the negligible performance improvement achieved by overlapping the blocks does not pay off the large increase of computation.

For each of this blocks we estimate \mathbf{C}_u by subtracting \mathbf{C}_w from the sample covariance of its coefficients, \mathbf{C}_y .

3) *Coefficient estimation*: For each neighborhood \mathbf{y} of observed coefficients, we estimate x_c , the reference coefficient, at the center of the neighborhood. The Bayes least squares (BLS) estimate can be expressed in this case [15] as

$$\mathbb{E}\{x_c|\mathbf{y}\} = \int_0^\infty p(z|\mathbf{y}) [\mathbb{E}\{\mathbf{x}|\mathbf{y}, z\}]_{x_c} dz.$$

Thus, the solution is written as the average of the Bayes least squares estimate of x_c when conditioned on z , weighted by the posterior density of z , $p(z|\mathbf{y})$. The advantage of the GSM model is that the coefficient neighborhood vector \mathbf{x} is Gaussian when conditioned on z . This fact, coupled with the assumption of additive Gaussian noise means that the expected value inside the integral of (2) is simply a Wiener estimate:

$$\mathbb{E}\{\mathbf{x}|\mathbf{y}, z\} = z\mathbf{C}_u(z\mathbf{C}_u + \mathbf{C}_w)^{-1}\mathbf{y}.$$

We can simplify the dependence of this expression on z by diagonalizing the matrix $z\mathbf{C}_u + \mathbf{C}_w$ (see [15]). Note that, unlike in [15] now this diagonalization depends on the spatial location (as \mathbf{C}_u is a function of (n, m)), and, thus, it needs to be computed once per each considered spatial location of each subband.

B. Deblurring-by-Denoising using space-variant GSMs

We face the problem of image restoration, where prior to noise addition, the signal has been degraded by (typically) a low-pass filter. The observation is modelled as follows:

$$y(n, m) = h(n, m) * x(n, m) + w_0(n, m). \quad (2)$$

where h is a convolution kernel (assumed known) applied to the original image x , usually called Point Spread Function, PSF. w_0 is Gaussian noise of known (but arbitrary) power spectral density. The same BLS-GSM local approach used in [15] for image denoising was used in [35] for image restoration. However, in this case, the fact of having a local description of image and degradation caused problems when trying to compensate for certain convolution kernels. Let's analyze this problem. Given a small neighborhood for the local model we can only aspire to compensate for smoothly varying filters in the frequency domain having no zeros (so their regularized inverse filters will also be smooth in frequency, and, as a consequence, small in the image domain). A fixed absolute neighborhood size, thus, would very severely limit the deconvolution power of the method. A better option is to use a fixed neighborhood size for all the scales in a multi-resolution representation. In such a way we are able to represent larger effective neighborhoods as we go up in the pyramid. We explain this with a 1D example in Fig 3. Above we show the Fourier representation, in magnitude, of a spatially limited signal, which corresponds to the same neighborhood size, measured in number of coefficients, at different scales. As we go up in the scale, the frequency support becomes smaller, and, thus, more rapidly varying responses are obtained. So, thanks to multi-resolution,

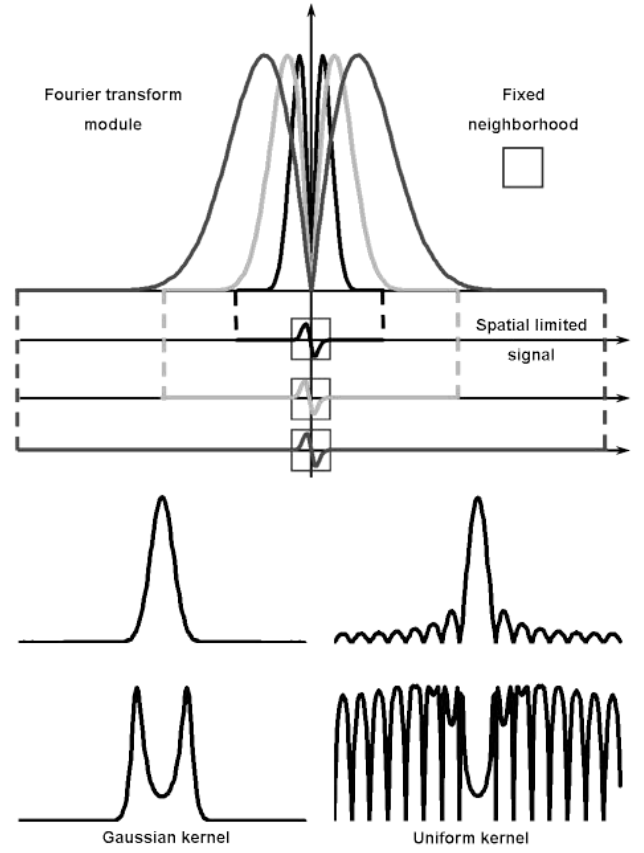


Fig. 3. Illustration of the restrictions on the shape in the frequency domain of the filters that we can represent using a local approach (fixed small neighborhood) with multi-resolution. The same spatial limited signal is shown at three different scales. This scheme allows us to represent (and to compensate for) slowly varying filter responses at high frequencies, with no zeros, and more rapidly varying frequency responses at low frequencies. At the bottom two typical 1D low-pass filters are shown for comparison, together with a regularized version of their inverse filters. We can see that the Gaussian kernel can be well compensated for using this kind of local multi-resolution approach, whereas the uniform kernel can not.

we can use (and compensate for, whenever there are no zeros) smooth filters at high frequencies, less smooth at medium frequencies, and rapidly varying at low frequencies. Such constraint is compatible with many low-pass kernels (e.g., Laplacian, Gaussian, etc., see bottom left of Fig. 3). However, there are some practically important kernels, such as those integrating uniformly in their spatial supports, whose behavior in frequency is not flat at all at high frequencies, but presenting strong oscillations around zero. Trying to compensate for those zeros requires abrupt responses for the regularized inverse filtering, which implies necessarily a large spatial support (see bottom right of Fig. 3). Therefore, uniform blurring cannot be properly treated by using this kind of local multi-resolution schemes. Another drawback of the referred method [35] is that, to work properly, it requires a reasonably accurate estimation of the cross-covariance among the original and the filtered image. This implies, in turn, the need of estimating the power spectrum of the original image (P_x) from the degraded observation, which is a difficult

problem. Using a general model for P_x translates in this case into a significant performance fall.

1) *Deblurring-by-Denoising*: In order to keep the advantages of using a local model, while overcoming its drawbacks, we use here the two-step scheme depicted in Fig. 1. This deconvolution strategy was proposed more than ten years ago [28], and it has been followed by many authors [29], [30], [31], [32], [33]. The two steps are: a) Blur compensation through global linear filtering, and, b) local denoising of the blur-compensated observation.

We can express the observation model of Eq. 2 in the frequency domain as

$$Y_0(u, v) = H(u, v)X(u, v) + W_0(u, v),$$

where⁴ H is the Fourier transform of the blur kernel h . Let G be the filter applied in step 1 (here in after, prefilter). In order to compute the optimal prefilter in MSE sense, G_{opt} , we must couple both steps:

$$G_{opt} = \arg \min_G \{\mathbb{E}\{\|X - f(GY_0)\|^2\}\}, \quad (3)$$

where f represents here the denoising operator (step 2), so $f(GY_0)$ represents the output of the whole process. The expectation inside Eq. 3 can be expressed as:

$$\begin{aligned} \mathbb{E}\{\|X - f(GY_0)\|^2\} &= \mathbb{E}\{\|(X - GHX) + (GHX - f(GY_0))\|^2\} \\ &\simeq \mathbb{E}\{\|(X - GHX)\|^2\} + \mathbb{E}\{\|(GHX - f(GY_0))\|^2\}, \end{aligned}$$

where the r.h.s. of the latter approximation represents the sum of the quadratic errors coming from the deblurring and the denoising steps. This assumes that both errors are approximately decorrelated, which is true, at least under usual conditions⁵. Thus, to compute the optimal prefilter G we need an explicit error model for the denoising step. To obtain such an error model for a non-linear adaptive denoising method is a major challenge, and, to the best of our knowledge, no one has proposed such a model to this date. In the absence of a globally optimal solution, next we follow a marginally optimal procedure which is, nevertheless, efficient and robust.

Step 1: Global blur compensation

This step provides a blur compensated image by applying a global filtering to the observed image. Instead of estimating an optimal filter we reduce the degrees of freedom from $I_x \times I_y$ (image and filter size) to only one, corresponding to the regularization parameter, α . We use, as in other works [29], [30], [33] a generalized Wiener filter:

$$G = \frac{H^*}{|H|^2 + \alpha \cdot \left(\frac{P_{w_0}}{P_x}\right)},$$

where P_{w_0} and P_x are the power spectral densities of noise and signal respectively. This regularized inversion compensates for the filter phase response. The balance between noise and blur suppression is controlled by the

⁴For notational simplicity from here on we drop the frequency indices (u,v).

⁵We have empirically tested this decorrelation through simulations

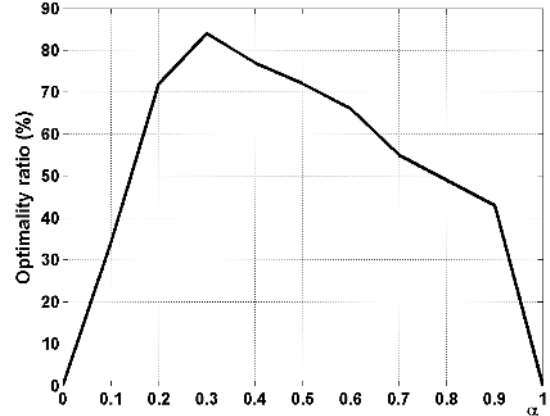


Fig. 4. Optimality ratio of the regularization parameter α . See text for details.

regularization parameter α . Low values yield in noisy estimations ($\alpha = 0$ is just an inverse filter⁶) whereas high values tend to suppress too much signal. The higher is α , the higher is the overall linear distortion, but the lower it is, the noisier is the estimation. $\alpha = 1$ yields the Wiener filter, which is a far too high value when using typical images and non-linear denoising methods in the second step, as we show in subsection III-B2.

Step 2: Local denoising

After step-1, the observation yields:

$$Y_1 = GY_0 = GHX + GW_0 = H_r X + W_1.$$

Because G has been optimized in the first step, no further attempt to correct the remaining linear distortion $H_r = GH$ is made, and what is left is to denoise the prefiltered image.

2) *Prefiltering parameter estimation*: Based on the typical images power spectrum [46], [47], we have used a simple model for P_x , namely $\hat{P}_x = k/f^2$, where f represents absolute frequency and k an arbitrary constant. In order to force the model to be consistent with the estimated original variance ($\hat{\sigma}_x^2 = \sigma_{y_0}^2 - \sigma_{w_0}^2$), the following expression must hold:

$$\hat{\sigma}_x^2 = \int_{f>0} |H(f)|^2 \hat{P}_x f df = \int_{f>0} |H(f)|^2 \frac{k}{f} df,$$

hence k is set to:

$$k = \hat{\sigma}_x^2 / \int_{f>0} |H(f)|^2 / f df.$$

Despite its simplicity, this model is effective and, in contrast to other methods (such as [35]), the associated P_x estimation error does not seem to have a critical impact on the overall performance.

Unlike other authors [28], [30] that used a model to choose the regularization parameter α , we have computed it empirically, training with a set of standard test images (*House*, *Cameraman*, *Barbara*, *Boat*, and *Lena*), using Gaussian convolution kernels and white Gaussian noise of 0.4 to 3.2

⁶We point out that it is critical for the well behavior of the filter G and for the final performance of the method to ensure that P_{w_0} is strictly positive on the whole frequency plane. This is a practical concern when P_{w_0} is estimated from real images.

and 1 to 16 standard deviations, respectively (in a geometric scale). For each combination of degradation parameters we have computed the optimal α through golden search. We also computed a quality decrease tolerance interval around the optimal value, according to a maximum performance decrease of 0.05 dB. Finally we chose the common α value providing the highest proportion of experiments within the quality tolerance intervals. Fig. 4 shows the optimality ratio for each alpha value, showing a peak for $\alpha = 0.3$, for which value around 84% of the experiments provide quasi-optimal results. We have also tested the method with very different PSFs [2] (such as a uniform kernel), and the maximum of the histogram was still very close to 0.3, although the percentage dropped slightly (73% for the uniform kernel). We have also tested that the results are fairly independent on the particular images chosen for the training, by comparing these results with those obtained by excluding each time the image to be processed in the training set.

IV. IMPLEMENTATION

A. Boundary and block handling

As demonstrated in [15], careful boundary handling is crucial for achieving a high denoising performance. In this case it is even more critical than for [15], because of the image partition into blocks. To illustrate this, let's consider a $L \times L$ GSM neighborhood in a N_s scales pyramid. In the upper part of Fig. 5 we can see the corner of a subband at the the highest scale in the pyramid ($L = 3$ and $N_s = 4$ in this example). All GSM neighborhood with reference coefficients at the boundaries of the subband are incomplete, lacking those coefficients marked by the dashed lines. To solve this, the observation is extended, by mirror reflection with no repetition, $2^{(N_s-1)}(L-1)/2$ pixels at each side, which yields the required extra coefficients for the subbands at the top of the pyramid (see dashed square in Fig. 5). When dealing with block-based methods, the boundary artifacts are potentially stronger, because of the larger proportion of boundary coefficients compared to the whole image. Same as before, each block must be extended $(L-1)/2$ coefficients at each side. The inner blocks are enlarged with true signal, whereas boundary blocks use the mirror extension made for the whole subband. This produces a small overlapping between extended blocks. We have tested that increasing this overlapping does not translate into any significant improvement.

Another practical issue is the block arrangement within subbands. In order to have complete blocks in all subbands we should ensure that all the subbands have dimensions that are integer multiples of the block size, i.e., $\mathbf{I}_{x,y}/(2^s B)$ is an integer vector for $s = 0 \dots N_s - 1$, being $\mathbf{I}_{x,y}$ the image dimension vector. As original image dimensions will not comply, in general, with this requirement (see dark shaded area in Fig. 5), they will generally require to be extended (by mirror reflection) to also fit this constraint (see light shaded areas in Fig. 5). In practice we apply a relaxed version of this criterion, processing the whole subbands when they are smaller than the chosen block size, so effectively reducing the range of s for imposing this constraint.

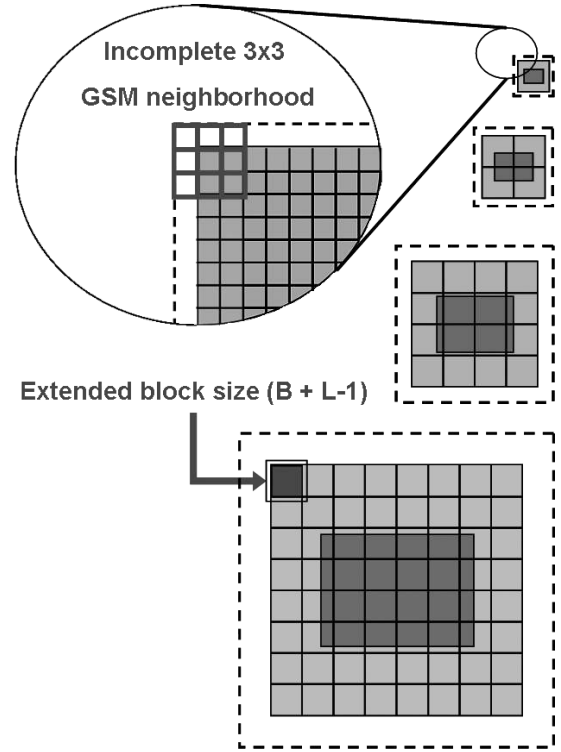


Fig. 5: Multiresolution block fitting.

B. Denoising experiments

We have used the translation-invariant Haar pyramid (TIHP) and the full steerable pyramid (FSP) as image representations. We chose the number of scales to 4 and 5 for 256×256 and 512×512 images respectively. Note that the number of orientations is fixed for the TIHP (3), whereas we used 8 orientations for the FSP (as in [15]). We used a 3×3 GSM neighborhood around each coefficient (same as in [15]). In case of FSP we also include a coefficient at the same location and orientation at the next coarser scale (a parent). Different block sizes B were tested for estimating the signal covariance ($B = 16, 32, 64, 128, F$, where F means the whole image). Both the GSM neighborhood and the block size, are the same in all pyramid levels, except for the subbands at the top of the pyramid, which have no parent, and those subbands with a size smaller than the block.

C. Deblurring experiments

We have implemented the global prefiltering (step 1) in the Fourier domain, using $\alpha = 0.3$. For the denoising (step 2) we have selected manually TIHP or FSP based on the relative amount of texture in the image: FSP was chosen for texture-rich and TIHP for poorly textured images. The number of orientations were kept the same as for denoising whereas the chosen number of scales was 4 for all images sizes. The GSM neighborhood was also 3×3 , but with no parent now in either representations. The block size was fixed to 32×32 .

D. Computational cost

As a guide, Table I shows the computation time (in seconds) of our unoptimized denoising and deblurring *Matlab*[®] 7.0

Representation	FSP (8 Orientations)		TIHP (3 Orientations)	
	Image size \Rightarrow	256 \times 256	512 \times 512	256 \times 256
<i>Algorithm</i>	16 \times 16			
Denoising	108.51	1090.07	46.18	451.29
Deblurring	107.00	1091.39	45.39	447.77
<i>Algorithm</i>	32 \times 32			
Denoising	39.85	312.55	16.40	132.25
Deblurring	39.13	312.48	16.40	133.20
<i>Algorithm</i>	64 \times 64			
Denoising	22.65	119.36	8.76	52.19
Deblurring	22.49	122.26	9.22	54.27
<i>Algorithm</i>	<i>Original BLS-GSM</i>			
Denoising	16.14	62.92	7.22	32.44
Deblurring	17.26	61.11	7.50	29.66

TABLE I. Averaged computation time in seconds (10 trials) comparing different images sizes, block sizes and representations for both methods. Four scales are used.

code, on *Windows XP* over a 3.0 GHz *Intel Pentium IV* CPU with 1Gb RAM, for different representations, blocks and images sizes. For comparison, we also include the computation time of the original BLS-GSM at the bottom part of the table. We extract three basic patterns: *a)* The cost scales approximately inversely with B , roughly a factor of 3 when we halve the block; *b)* The cost of the first step in the deblurring algorithm is negligible when compared to the cost of the second step; and *c)* The computation time is significantly higher, but still comparable to that of original BLS-GSM. The main difference (in cost terms) between the original BLS-GSM and the proposed method is the increase of calls to functions. The proposed method usually makes one call per *block*, whereas the original BLS-GSM makes only one per *subband*. Note also that the number of calls are multiplied by a factor of 4 when we halve the block size. This has a big impact due to using unoptimized interpret code in our current implementation. Finally, the cost scales roughly with a factor 8 when we double the image size.

V. RESULTS AND DISCUSSION

We have used five standard test images to evaluate our algorithms: *Barbara*, *Lena*, *House*, *Cameraman* and *Peppers*. We have chosen *Barbara* and *House* for all visual comparisons, as two examples with very different amount of texture.

A. Denoising

We selected four standard images: *Barbara*, *Lena* (both 512 \times 512), *House* and *Peppers* (both 256 \times 256). We have degraded them by adding additive white Gaussian noise of different standard deviations ($\sigma = 5, 10, 15, 20, 25, 50$).

1) *Model Features:* In Table II we present the results of our proposed algorithm, in terms of peak signal-to-noise ratio (PSNR), $10 \log_{10}(255^2/\sigma_e^2)$ in dB, where σ_e^2 is the mean square error. We include results for the representations used, TIHP and FSP, and compare our method (column labelled Two-Level Adaptive, 2LA) with the standard version of BLS-GSM [15] (labelled One-Level Adaptive, 1LA). We also include the block size, B , for which the method provided the best results. In Fig. 6 we show some visual results (cropped and zoomed, for artifacts visibility) on *Barbara* and *House*

σ_w / PSNR	FSP		TIHP		Block Size
	1LA	2LA	1LA	2LA	
<i>Lena</i>					
5/ 34.15	38.49	38.55	38.19	38.34	32
10/ 28.13	35.60	35.66	35.23	35.38	64
15/ 24.61	33.90	33.96	33.50	33.65	64
20/ 22.11	32.67	32.71	32.25	32.40	64
25/ 20.17	31.69	31.72	31.26	31.40	64
50/ 14.15	28.61	28.61	28.21	28.21	F
<i>Barbara</i>					
5/ 34.15	37.78	38.19	37.19	37.84	16
10/ 28.13	34.02	34.60	33.13	34.13	16
15/ 24.61	31.83	32.49	30.76	31.85	16
20/ 22.11	30.27	30.94	29.08	30.19	16
25/ 20.17	29.07	29.76	27.81	28.81	32
50/ 14.15	25.42	26.04	24.33	25.03	32
<i>House</i>					
5/ 34.15	38.65	38.88	38.23	38.55	16
10/ 28.13	35.29	35.44	35.32	35.56	16
15/ 24.61	33.54	33.64	33.73	33.94	32
20/ 22.11	32.30	32.34	32.54	32.71	32
25/ 20.17	31.32	31.30	31.59	31.71	32
50/ 14.15	28.22	28.18	28.36	28.38	128
<i>Peppers</i>					
5/ 34.15	37.30	37.43	37.57	37.72	16
10/ 28.13	33.74	33.81	34.06	34.24	16
15/ 24.61	31.7	31.72	32.02	32.18	16
20/ 22.11	30.27	30.24	30.57	30.67	16
25/ 20.17	29.17	29.10	29.45	29.50	16
50/ 14.15	25.94	25.94	26.11	26.11	F

TABLE II. Denoising performance expressed as Peak Signal-to-Noise Ratio in dB. First column shows the noise standard deviation and the corresponding PSNR. Next four columns correspond to FSP and TIHP representations in the one-level adaptive (1LA) and two-level adaptive (2LA) version (highlighted the best of those 4 results). Last column shows the block size used for the 2LA method, where F means the full image.

images. We include results for TIHP and FSP (second and third rows) and a comparison with standard BLS-GSM (by columns). We now present a discussion along the followings directions:

Comparison to standard BLS-GSM

The enhancement of the model proposed in section II-B, provides us a general improvement of the results over standard BLS-GSM method, both visually and in MSE sense. Comparing the columns 1LA-2LA of Table II we can see a significant performance increase. The new results are better (with no exception) for both representation TIHP and FSP. As we can see, the improvement is more important for images having local spectral peaks (like *Barbara*), whereas for other images (e.g. *Lena*) the improvement is more modest. This behavior is related with the amount of covariance variability among different regions in the image (see Fig. 2 for details). We can compare visually the differences between our method and the standard BLS-GSM, for *Barbara* and *House* images, in the columns of Fig. 6. Note how artifacts on the roof line of *House* disappear with the new method (second column). For *Barbara* we can see a very strong improvement, the new results having much less artifacts, and much more texture is

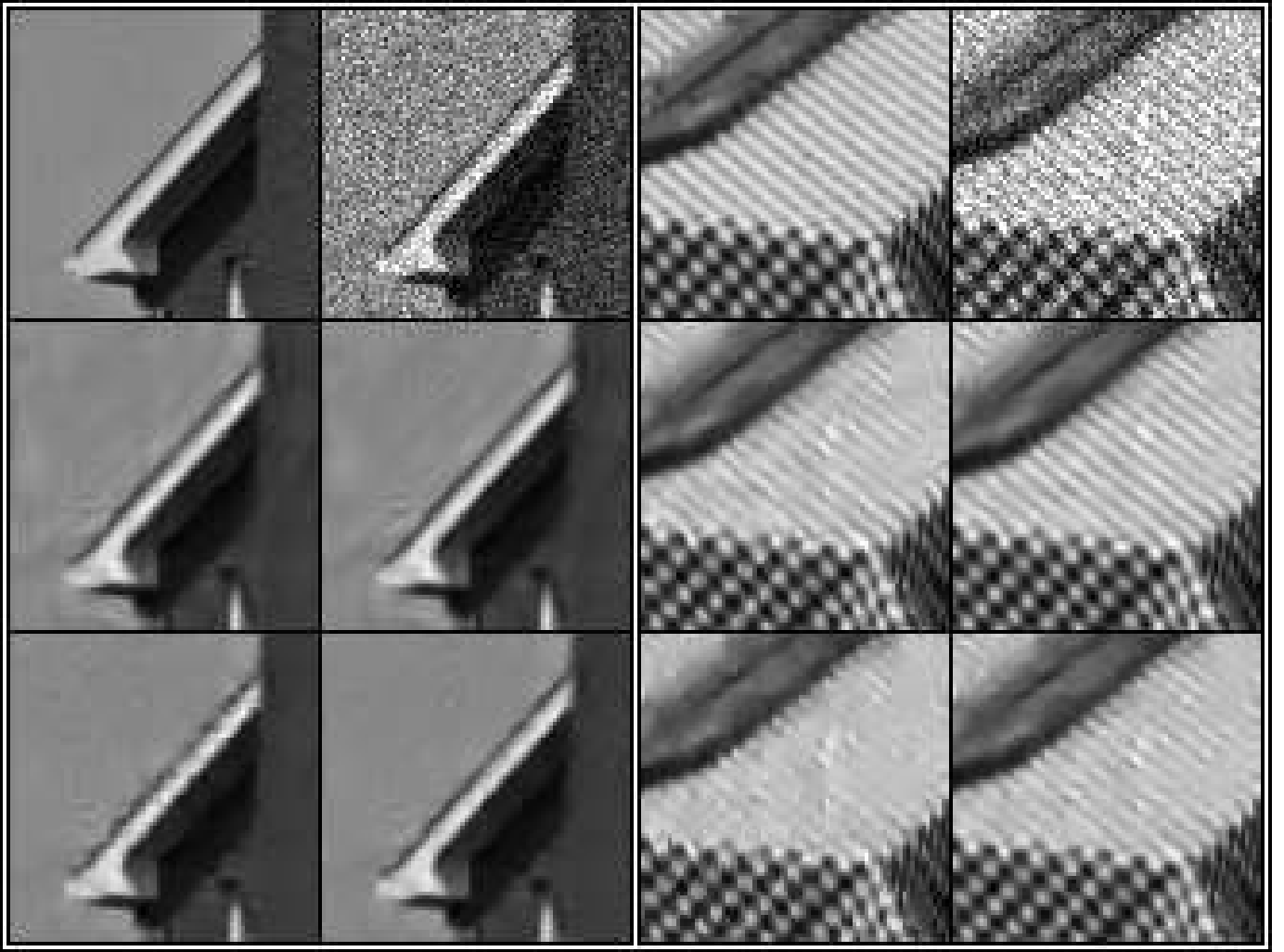


Fig. 6. Visual comparison results on *House* and *Barbara* images cropped to 80×80 pixels (starting in pixel [1,78] and [1,221] respectively, [column, row]). We used 32×32 and 16×16 block sizes for *House* and *Barbara*, respectively. From left to right and top to bottom (House / Barbara PSNR in dB): Original; Noisy ($\sigma_w = 20$), (22.11/22.11); BLS-GSM [15] with FSP (32.30/30.27); Our method with FSP (32.34/30.94); BLS-GSM [15] with TIHP (32.54/29.08); Our method with TIHP (32.71/30.19).

recovered.

Representation comparison: FSP vs. TIHP

As we can see in Table II, the results confirm what we pointed out in section II-A: Different representations fit better different images, depending on their contents. *House* and *Peppers*, two poorly textured images, are best denoised using the TIHP, whereas for *Lena* and *Barbara* the best results are achieved with FSP. We can also see the significant difference in visual terms, by inspecting *House* and *Barbara* images in the second (FSP) and third (TIHP) rows of Fig. 6. In the case of *House*, the visual difference is significant (less artifacts in general), but not dramatic. But in the case of *Barbara* the difference is striking: FSP provides a much better result than the one obtained with TIHP. If one of them should be chosen basing only on results (ignoring computational cost), under a minimax risk criterion, FSP would be the choice.

Another issue is the different relative gains between the new method when using TIHP vs. FSP, with respect to the standard BLS-GSM. By comparing in Table II the performance difference between our method and the original

BLS-GSM for each image and noise level (by rows), we see that the improvement is bigger when we use the TIHP. The main reason is that whereas FSP kernels are highly selective in frequency, TIHP kernels are not, and hence there is more spatial variability of the covariance inside each subband in the later representation. This increases the difference between local covariances and the global one (see details in Section II-B). We have also tested the benefit of averaging FSP and TIHP results (in the same way as other authors [42]), obtaining an average increase of 0.15 dB over the best one.

Block size considerations

The optimal block size depends on the image contents in the first place, but also on the noise level. The lower the signal-to-noise ratio is, the larger becomes the optimal size. This makes sense, as local signal covariance estimation stops being reliable for small block sizes in the presence of high noise, so covariance spatial adaptation no longer pays off. In Fig. 7 we show, for *Barbara* and *House*, a plot of PSNR (with range adjusted between 0 and 1) when using $B = 16, 32$ and 128 vs. the noise standard deviation. Here one corresponds to the

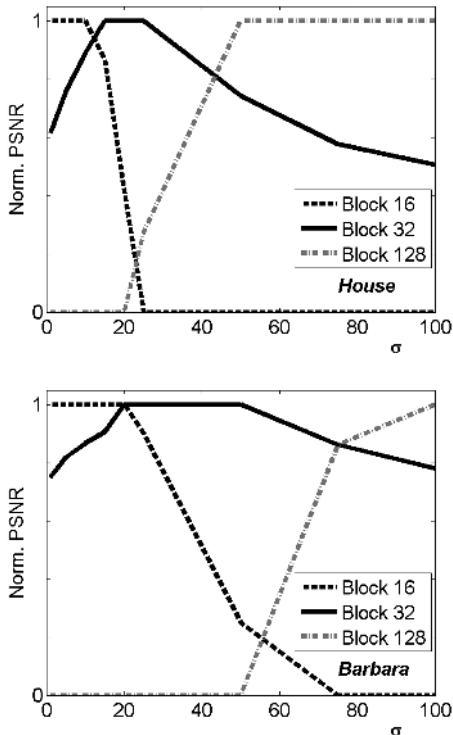


Fig. 7. PSNR vs. noise standard deviation for *House* and *Barbara* images (adjusted between 0 and 1). Dashed black line 16 \times 16 block, solid line for 32 \times 32 block and dot-dashed gray one for 128 \times 128 one. Note that the "optimal" block size increases with the noise level.

best result obtained with the set of block sizes tested, and zero to the worst result. We can state, as expected, that the best result for low noise levels are achieved with the smallest size block ($B = 16$), whereas $B = 128$ provides better performance for high noise variance. Overall, it seems that a finer adaptation ($B = 16$) provides better results than coarser covariance adaptations. However, differences are, in general, small, and, from a computational cost view, the smaller is the block the higher becomes CPU time. As we can see in Fig. 7, the curve associated to a block size of 32 is a good trade-off in practice, providing best results for moderate noise levels and not dramatically decreasing when in the extremes.

2) *Comparison to state-of-the-art methods:* In Fig. 8 we show a visual comparison of our method and, what to the best of our knowledge, are the three ones with highest PSNR performance available in the literature [22], [21], [23]. For *House* our method behaves the worst in PSNR terms, mainly because the other methods recover the wall texture and ours not. Its visual appearance is, nevertheless, convincing. For instance, the roof line is sharper and less contaminated by artifacts in our result than in [22]. For *Barbara* our method behaves similarly to the other three, both visually and in MSE terms. Fig. 9 shows, also for *Barbara* and *House*, a graphical comparison, in terms of input - output PSNR, among our method and the top five reporting the best results [21], [22], [19], [20], [23]. Our method is represented by a solid line, whereas the standard BLS-GSM is included as a reference using a dash line. As we can see, our method behaves competitively to current state-of-the-art [21], especially considering its

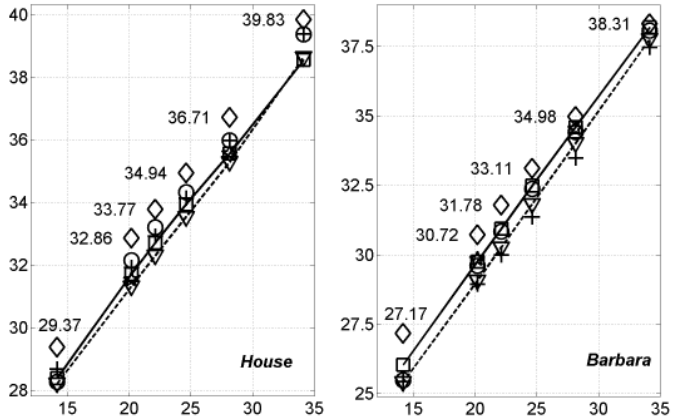


Fig. 9. PSNR output vs. PSNR input, in dB. We use TIHP for *House* and FSP for *Barbara*. Our results, for $\sigma_w = \{5, 10, 15, 20, 25, 50\}$, are in Table II. We compare to three state-of-the-art methods: Diamonds for [21]; Crosses for [22]; Circles for [23]; Down triangles for [15] and Squares for our new method. The latter two results are represented with dashed and solid lines respectively.

	σ	NOISE				
		1	2	4	8	16
B	0.4	2.86	1.86	2.81	5.03	7.63
L	0.8	6.65	4.84	3.73	3.92	5.85
U	1.6	3.87	3.18	2.72	3.01	4.73
R	3.2	3.21	2.85	2.56	2.66	3.87

TABLE III. Averaged results as Increment of Signal-to-Noise Ratio (ISNR), in dB, of the training data set (see sec. III-B2 for details).

moderate computational cost. We also refer the reader to [53] for a recent perceptual comparison of state-of-the-art denoising methods which ranked a preliminary version of this method among the top three.

B. Deblurring

In Table III we show averaged Increment in Signal-to-Noise Ratio (ISNR) results of applying our deblurring method to the training data set (see Section III-B2 for details). As we can see, the results exhibit a significant improvement for the whole range of tested blur and noise degradations. In Table IV we show the results for six deconvolutions experiments with *House*, *Cameraman* (using THP) and *Barbara* (using FSP) comparing our method with the previous BLS-GSM restoration [35], and with those which, to the best of our knowledge, represent the current state-of-the-art [33], [24]. In the first two columns we have replicated the experiments from [32]. The kernel used (PSF1) is $h_{i,j} = (1 + i^2 + j^2)^{-1}$, for $i, j = -7 \dots 7$. The next column reproduces the experiment in [28], [30], which uses a 9×9 uniform kernel (PSF2). The fourth column uses a 5×5 separable kernel (PSF3) with coefficients $[1, 4, 6, 4, 1]/16$ as in [54]. The last two columns use Gaussian kernels with 1.6 and 0.4 standard deviation (PSF4 and PSF5 respectively). The noise standard deviations are shown in the second row of the table. The results show that [33] provides the best results in a 44.4% of the experiments, [24]+[25] in a 16.7% and the proposed

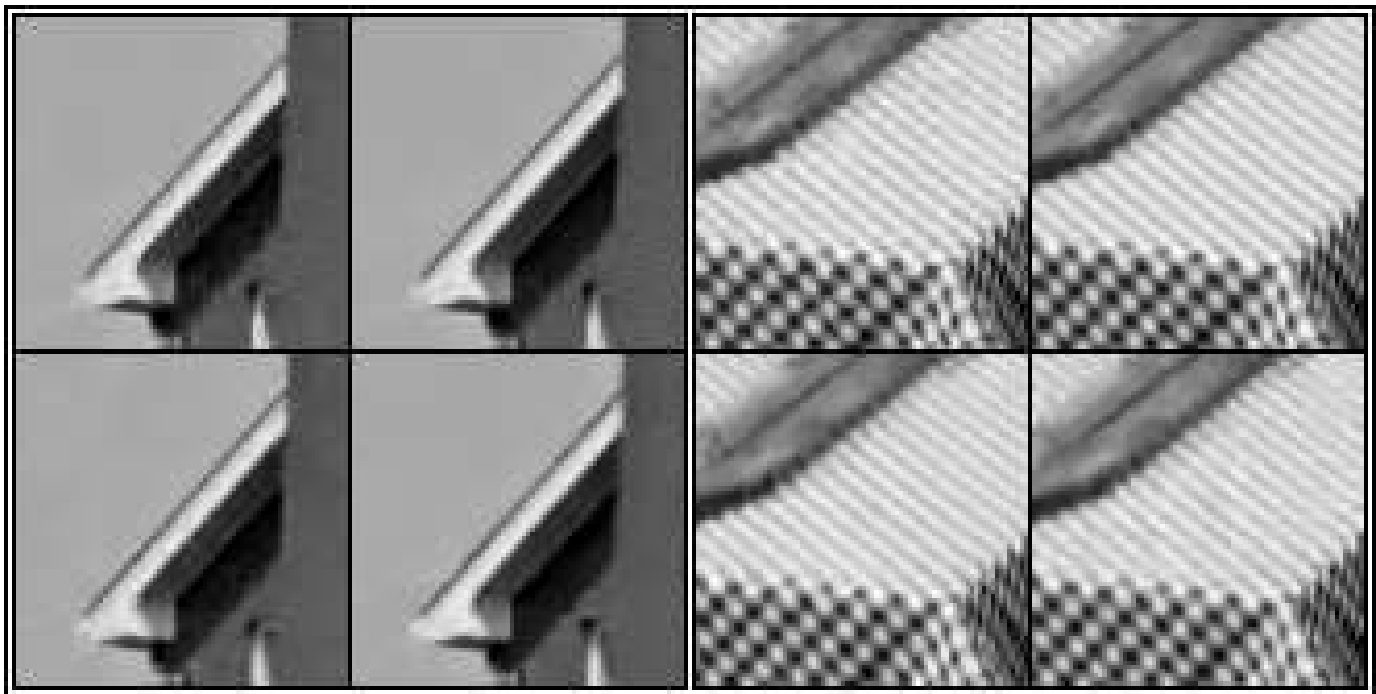


Fig. 8. Visual comparison results on *House* and *Barbara* images cropped to 80×80 pixels (starting in pixel [1,78] and [1,221] respectively, [column, row]). Initial PSNR 28.13 dB in both cases ($\sigma_w = 10$). From left to right and top to bottom (*House* / *Barbara* PSNR): Results for [22] (35.98/33.48); Results for [21] (36.71/34.98); Results for [23] (35.94/34.42); Our results using a 16×16 block with TIHP/FSP representation (35.56/34.60).

Blur	<i>PSF 1</i>	<i>PSF 2</i>	<i>PSF 3</i>	<i>PSF 4</i>	<i>PSF 5</i>	
$\sigma_w \Rightarrow$	$\sqrt{2}$	$\sqrt{8}$	$\sqrt{0.308}$	7	2	8
<i>HOUSE</i>						
PSNR	25.62	25.47	24.11	28.10	27.83	30.01
[24]	8.47	6.63	10.71	4.22	4.49	4.76
[35]	8.46	6.93	-0.44	4.37	4.34	5.98
[33]	9.05	7.64	10.71	5.10	4.03	6.13
Ours	8.64	7.03	9.04	4.30	4.11	6.02
<i>CAMERAMAN</i>						
PSNR	22.23	22.16	20.77	24.63	23.36	29.83
[24]	7.46	5.24	8.16	2.84	3.18	3.65
[25]	6.93	4.88	7.59	2.94	-/-	-/-
[26]	7.40	5.15	8.10	2.85	-/-	-/-
[35]	6.84	5.29	-1.61	2.56	2.83	3.81
[33]	8.25	6.34	8.57	2.56	3.05	5.15
Ours	7.45	5.55	7.33	2.73	3.25	4.19
<i>BARBARA</i>						
PSNR	23.33	23.26	22.49	24.21	23.77	29.77
[24]	3.76	1.99	3.98	0.9	0.92	2.55
[35]	5.70	3.28	-0.27	1.44	0.95	4.91
[33]	5.73	3.01	4.88	1.58	0.91	4.04
Ours	6.85	3.80	5.07	1.94	1.36	5.27

TABLE IV. Performance comparison with other state of the art methods in terms of Increment of Signal-to-Noise ratio (ISNR), in dB. First row shows the blurring kernel used (see text for details), and the second denotes noise standard deviation. Best results are highlighted.

method in a 38.9% of the experiments. Comparing with the BLS-GSM restoration original method [35], the new method provides a very significant improvement for all experiments, especially for the uniform degradation (PSF2), where the local BLS-GSM approach completely fails (due to reasons explained in section III-B). Compared to [33], note that, in general, our method achieves better results with *Barbara*, and worse with

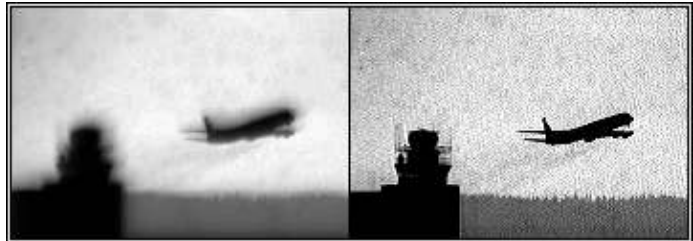


Fig. 11. Result of the method applied to a real degradation.

House and *Cameraman*. Being [33] a two-step method with a similar first global deconvolution step, the different behavior mainly comes from the denoising step [22] (based on the SA-DCT transform). This method is specially fitted to preserve sharp edges and contours (present in *House* and *Cameraman*), whereas it is not as powerful as ours to characterize local spectra. This translates into a poorer performance when dealing with highly texturized images (as *Barbara*).

In Fig. 10 we show a visual comparison among our method, the state-of-the-art methods [24], [33] and BLS-GSM standard restoration [35]. For *Barbara*, the degradation used is PSF1 with $\sigma = \sqrt{2}$. In this case our method provides the best result, both visually and in PSNR terms. Particularly, the diagonal lines of the table cloth are better recovered by our method than by our competitors. For the *House* image we have used PSF2 (uniform kernel) with $\sigma = \sqrt{0.308}$. Here we can see how the previous BLS-GSM restoration method [35] completely fails, whereas the new version provides results competitive to the state-of-the-art [24], [33].

Finally, we also tried our deblurring algorithm with real degradations. We restored a micro-display image captured

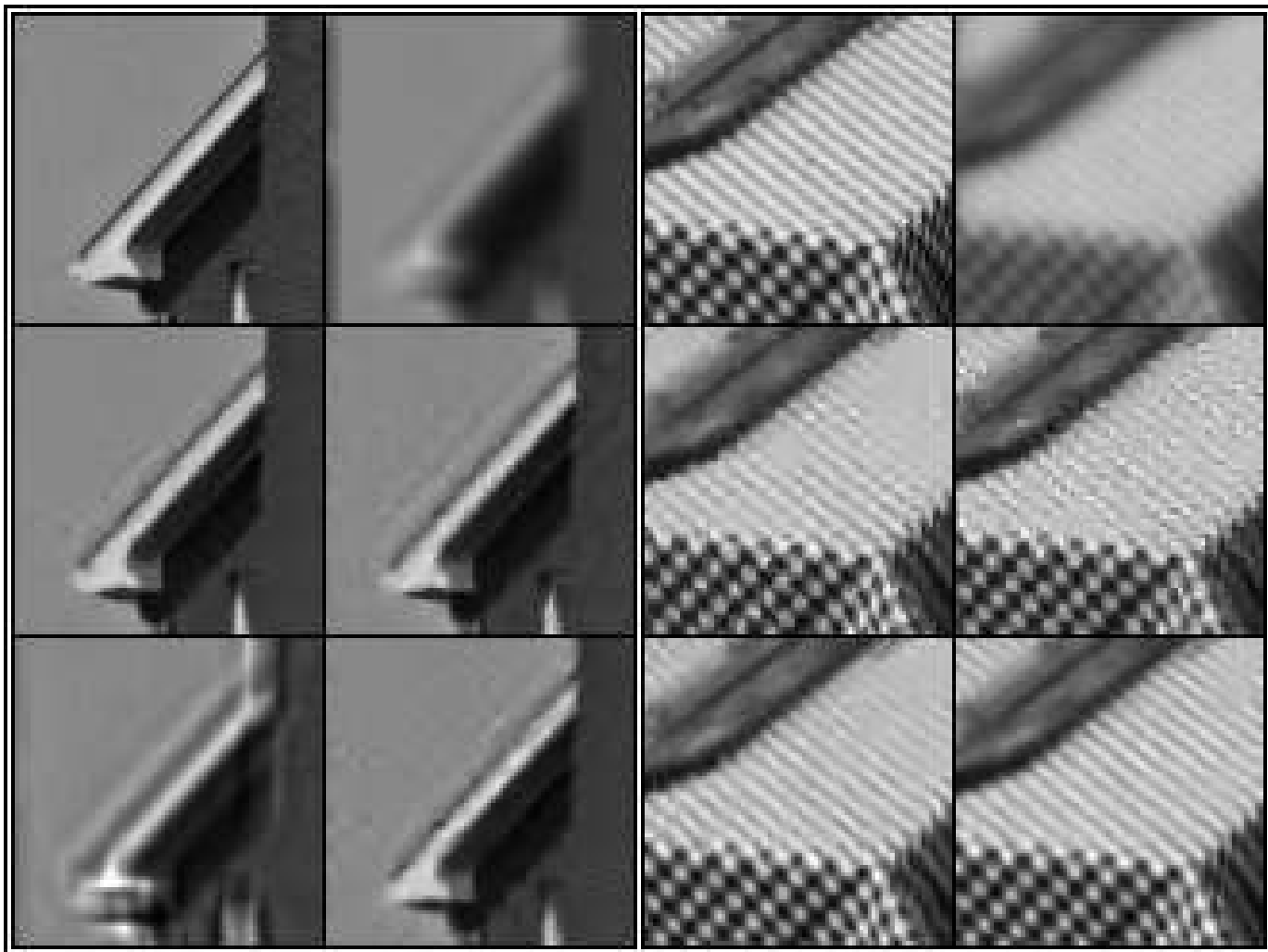


Fig. 10. Visual comparison results on *House* and *Barbara* images cropped to 80×80 pixels (starting in pixel [1,78] and [1,221] respectively, [column, row]). We include PSNR values, in dB (in agreement with ISNR values of Table IV). From left to right and top to bottom (*House* / *Barbara*): Original; Degraded image (24.11 / 23.33) using PSF2 & $\sigma_w = \sqrt{0.308}$ for *House* and PSF1 & $\sigma_w = \sqrt{2}$ for *Barbara* ; Results for [33] (34.76/29.05); Results for [24] (34.76/27.09); Results of standard BLS-GSM [35] with FSP (23.68/29.02); Our new method (*House* with TIHP, *Barbara* with FSP) (33.11/30.17).

under controlled laboratory conditions. We used a tilted lens to produce severe blur and the noise was reduced before processing by averaging 30 captures. In Figure 11 we show the degraded image and the result of our method using an estimation of the degradation parameters. As we can see the method behaves very well, substantially improving the visible detail in the image.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have presented two methods, for image denoising and deblurring. The denoising algorithm, based on space variant Gaussian scale mixtures, is applicable to any overcomplete *wavelet* representation. Although the proposed model is just a conceptually simple extension of the one presented in [15], it provides a significant improvement in performance terms, still with a moderate computational cost. We have shown that the reason of the improvement is that, by modelling the local covariance spatial variability, it provides a more realistic model and a higher noise filtering selectivity. The improvement depends on both the image content and

the representation used. We have tested two complementary representations concluding that one of them is more suitable for textured images (FSP), whereas the other is better for processing images for which texture is not so relevant (TIHP).

Our deblurring method is based on a 2-step decoupling scheme: linear global blur compensation and non-linear local non-white noise removal. This strategy provides flexibility to use any denoising method able to deal with non-white Gaussian noise, as the one presented here. The resulting algorithm overcomes the severe limitations imposed by local approaches to image deconvolution while keeping their advantages. In addition, and in contrast to some state-of-the-art methods, it provides a robust and simple solution to the problem of modelling and estimating the power spectral of the original image, avoiding the necessity of estimating it from a first estimation of the original image (*re-estimation*).

As shown, the proposed empirical link between the two steps of the method (parameter setup) is suboptimal. It would have been much more elegant to obtain the absolute optimal prefilter by using an error model of the denoising method.

Unfortunately, that is a difficult goal, given the non-linear character of high performance denoising methods. Nevertheless, the proposed empirical approach for choosing the prefiltering parameters is simple and robust: we have tested that the same prefiltering parameters can be used for different degradations without provoking a serious performance fall. This strategy is also very flexible, because the parameters can be easily tuned to any kind of degradation by training with kernels of the desired shape. All these features, coupled with the moderated (and image-content independent) computational cost, make this method suitable to restore images in real situations [55].

Summarizing, the solutions presented here are model-based, with few ad-hoc components that do not compromise a general conceptual coherence. Despite their simplicity, they provide state-of-the-art performance, both in denoising (slightly worse results than the best published ones) and deblurring (similar to the best ones on average).

A conceptually simple extension of the application of the methods presented here is to associate each spatial location with a different blurring kernel and/or with a different noise power spectral density. That would allow us to restore images with spatially variant degradations. This is interesting when dealing with low-quality optics, which can not be accurately modelled as convolutions, because they are not translation invariant. Although this implies some technical challenges (mainly related to block's boundary handling) we believe that it is a feasible goal, with a potential impact for the imaging industry.

We believe that BLS-GSM based models still have a strong potential for image processing. This work can be seen as a first exploration stage, with promising results, towards a more ambitious goal, namely the generalization of the GSM model to include more than one covariance matrix simultaneously for each subband. This new model, a "mixture of Gaussian scale mixtures", should be able to adapt to different textures in different spatial positions with no spatial constraints.

ACKNOWLEDGMENT

The authors would like to thank Profs. M. Elad, A. Foi, K. Dabov and M. Figueiredo for providing us with their results and software for testing. We also thank C. Dorronsoro and CIDA (Spain), for facilitating us the capturing of real degraded images in its laboratories.

APPENDIX A

TRANSLATION INVARIANT HAAR PYRAMID

We present an overcomplete version of the Haar orthogonal wavelet [45], which is translation invariant, due to the lack of aliasing, and which provides almost perfect reconstruction. Basically it is obtained by decimating its *à trous* [56] subbands from the third scale on, counting upwards. For simplicity, we start with the 1D case, and after that, the 2D extension is described.

In the top part of Table V are the descriptions (in spatial and frequency domains) of the 1D representation building functions. Note how the samples at the origin are underlined.

ONE DIMENSIONAL ANALYSIS FILTERS	
<i>Building functions</i>	
Spatial Domain	Fourier Domain
$\phi_1 = [\underline{1}, 1]/2$	$\Phi_1(\omega) = (1 + e^{-j2\pi\omega})/2$
$\psi_1 = [\underline{1}, -1]/2$	$\Psi_1(\omega) = (1 - e^{-j2\pi\omega})/2$
$\phi_2 = [1, \underline{0}, 1]/2$	$\Phi_2(\omega) = \cos(2\pi\omega)$
$\psi_2 = [1, \underline{0}, -1]/2$	$\Psi_2(\omega) = j \sin(2\pi\omega)$
ONE DIMENSIONAL RESPONSES	
First scale	Second scale
$H_1(\omega) = \Psi_1(\omega)$	$H_2(\omega) = \Phi_1(\omega)\Psi_2(\omega)$
$L_1(\omega) = \Phi_1(\omega)$	$L_2(\omega) = \Phi_1(\omega)\Phi_2(\omega)$
n^{th} scale	
$H_n(\omega) = L_{n-1}(\omega)j \sin(2^{n-1}\pi\omega), (\omega \leq 2^{1-n})$	
$L_n(\omega) = L_{n-1}(\omega) \cos(2^{n-1}\pi\omega), (\omega \leq 2^{1-n})$	
TIHP ANALYSIS FILTERS	
$A_v(u, v) = \Phi_1(u)\Psi_1(v)$	$B_v(u, v) = \Phi_2(u)\Psi_2(v)$
$A_h(u, v) = \Psi_1(u)\Phi_1(v)$	$B_h(u, v) = \Psi_2(u)\Phi_2(v)$
$A_d(u, v) = \Psi_1(u)\Psi_1(v)$	$B_d(u, v) = \Psi_2(u)\Psi_2(v)$
$A_l(u, v) = \Phi_1(u)\Phi_1(v)$	$B_l(u, v) = \Phi_2(u)\Phi_2(v)$

TABLE V. Top: One dimensional building functions, in spatial and frequency domain. The underlined coefficient represents the one at the origin. Middle: Description of the 1D representation responses. Bottom: TIHP analysis filters in terms of external products of the 1D building functions. The synthesis filters are their complex conjugates.

The middle part of Table V shows the descriptions of the scale responses in terms of the building functions. In Fig. 12 it is shown the responses (module) of a 4 scale case. The first two scales (two upper panels) coincide with those of the *à trous* representation (full resolution, no decimation). Further scales are decimated each time by a factor of 2, after applying an ideal low-pass filters to avoid intra-subband aliasing, as shown in the bottom half of Fig. 12. Dashed lines depict the corresponding *à trous* response. The vertical dashed lines indicate the cut-off frequency of the ideal low pass filter. Frequencies on the right side are removed, thus perfect reconstruction is no longer possible. This is not as damaging as it may seem, because image content is usually concentrated at low frequencies (e.g., we obtain reconstruction with PSNR of 59.59 dB for *Lena* and 44.60 dB for a 512×512 white uniform noise in $[0 \ 255]$, both with 5 scales). Also, note that the cut-off frequencies coincide with a zero filter response in all cases, and also with a zero derivative for the high-pass filters. Thus, discontinuities are avoided and so they are most of the artifacts caused by them (ringing, poor spatial localization, etc.).

The extension to 2D is done following the system diagram depicted in Fig. 13. The filters ($A(u, v)$ and $B(u, v)$ boxes) are built by properly doing external products of the 1D filters, as shown at the bottom of Table V. The pyramid is implemented by recursively splitting the image into a set of oriented subbands (vertical, horizontal and double diagonal) and a low pass band, which is subsampled by a factor of two along both axes. The reconstruction is achieved hierarchically by convolving each subband with its complex-conjugated filter, and adding the outputs.

REFERENCES

- [1] J A Guerrero-Colon and J Portilla, "Two-level adaptive denoising using Gaussian scale mixtures in overcomplete oriented pyramids," in *IEEE*

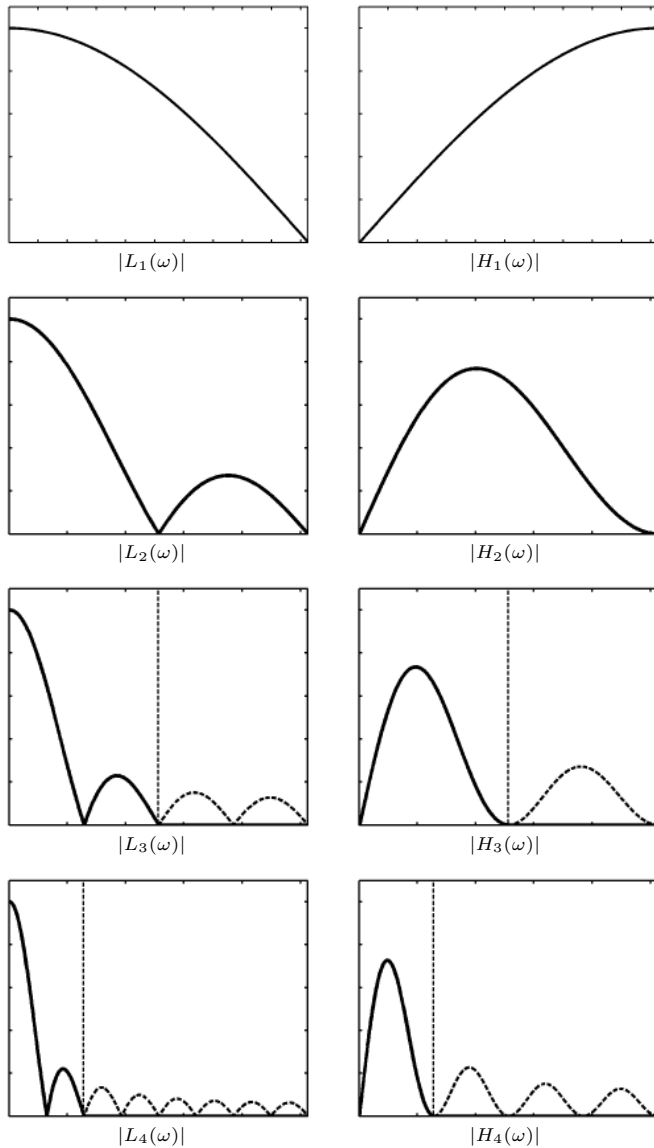


Fig. 12. Frequency response in module of the low and high/band pass filters in a 4 scale case. Dashed curves show the associated \hat{a} trous response. Vertical straight lines depict the cutoff frequency at each scale. See middle part of Table V for filters descriptions.

Int'l Conf. on Image Proc. Genoa, Italy, September 2005, vol. I, pp. 105–108.

- [2] J A Guerrero-Colon and J Portilla, “Deblurring-by-denoising using spatially adaptive gaussian scale mixtures in overcomplete pyramids,” in *IEEE Int'l Conf. on Image Proc.* Atlanta, GA, October 2006, pp. 625–628.
- [3] S Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” *IEEE Pat. Anal. Mach. Intell.*, vol. 11, pp. 674–693, July 1989.
- [4] J S Lee, “Digital image enhancement and noise filtering by use of local statistics,” *IEEE Pat. Anal. Mach. Intell.*, vol. PAMI-2, pp. 165–168, March 1980.
- [5] D L Donoho and I M Johnstone, “Ideal spatial adaptation via wavelet shrinkage,” *Biometrika*, vol. 81, pp. 425–455, 1994.
- [6] S G Chang, B Yu, and M Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image denoising,” in *IEEE Int'l Conf. on Image Proc.*, Chicago, October 1998, vol. 1, pp. 535–539.
- [7] A Pižurica, W Philips, I Lemahieu, and M Achery, “A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising,” *IEEE Trans. Image Proc.*, vol. 11, no. 5, pp. 545–557, May 2002.
- [8] L Şendur and I W Selesnick, “Bivariate shrinkage functions for wavelet-

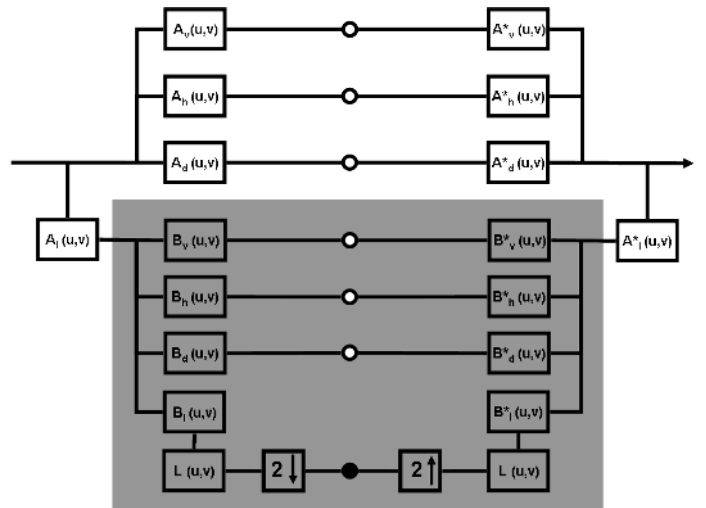


Fig. 13. System diagram for the translation invariant Haar pyramid. The input image is first split into six full resolution subbands by convolving with A and B filters. The recursion consists of inserting the diagram contents of the shaded region at the lowpass branch (solid circle).

based denoising exploiting interscale dependency,” *IEEE Trans. Signal Proc.*, vol. 50, no. 11, pp. 2744–2756, November 2002.

- [9] M K Mihçak, I Kozintsev, K Ramchandran, and P Moulin, “Low-complexity image denoising based on statistical modeling of wavelet coefficients,” *IEEE Trans. Signal Proc.*, vol. 6, no. 12, pp. 300–303, December 1999.
- [10] X Li and M T Orchard, “Spatially adaptive image denoising under overcomplete expansion,” in *IEEE Int'l Conf. on Image Proc.*, Vancouver, September 2000, vol. 3, pp. 300–303.
- [11] J Portilla, V Strela, M J Wainwright, and E P Simoncelli, “Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain,” in *IEEE Int'l Conf. on Image Proc.*, Thessaloniki, Greece, October 2001, pp. 37–40.
- [12] D Andrews and C Mallows, “Scale mixtures of normal distributions,” *J. Royal Stat. Soc.*, vol. 36, pp. 99–102, 1974.
- [13] M J Wainwright, “Visual adaptation as optimal information transmission,” *Vision Research*, vol. 39, pp. 3960–3974, 1999.
- [14] M J Wainwright and E P Simoncelli, “Scale mixtures of Gaussians and the statistics of natural images,” in *Adv. Neural Information Processing Systems*, Cambridge, MA, May 2000, vol. 12, pp. 855–861, MIT Press.
- [15] J Portilla, V Strela, M J Wainwright, and E P Simoncelli, “Image denoising using scale mixtures of Gaussians in the wavelet domain,” *IEEE Trans. Image Proc.*, vol. 12, pp. 1338–1351, November 2003.
- [16] S Mallat, G Papanicolaou, and Z Zhang, “Adaptive covariance estimation of locally stationary processes,” *Ann. Stat.*, vol. 26, no. 1, pp. 1–47, 1998.
- [17] D K Hammond and E P Simoncelli, “Image denoising with an orientation-adaptive Gaussian scale mixture model,” in *IEEE Int'l Conf. on Image Proc.*, Atlanta, GA, October 2006.
- [18] A Buades, B Coll, and J M Morel, “A non-local algorithm for image denoising,” in *IEEE Conf. on Comp. Vision and Pattern Rec.*, 2005, vol. 2, pp. 60–65.
- [19] C Kervrann and J Boulanger, “Optimal spatial adaptation for patch-based image denoising,” *IEEE Trans. Image Proc.*, vol. 15, no. 10, pp. 2866–2878, October 2006.
- [20] K Hirakawa and T W Parks, “Image denoising using total least squares,” *IEEE Trans. Image Proc.*, vol. 15, no. 9, pp. 2730–2742, September 2006.
- [21] K Dabov, A Foi, V Katkovnik, and K Egiazarian, “Image denoising by sparse 3-d transform-domain collaborative filtering,” *IEEE Trans. Image Proc.*, vol. 16, no. 8, pp. 2080–2095, August 2007.
- [22] A Foi, V Katkovnik, and K Egiazarian, “Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images,” *IEEE Trans. Image Proc.*, vol. 16, no. 5, pp. 1395–1411, May 2007.
- [23] M Elad and M Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Trans. Image Proc.*, vol. 15, no. 12, pp. 3736–3745, December 2006.
- [24] M Figueiredo and R Nowak, “A bound optimization approach to

- wavelet-based image deconvolution,” in *IEEE Int'l Conf. on Image Proc.*, 2005, vol. 2, pp. 782–785.
- [25] M Figueiredo and R Nowak, “An EM algorithm for wavelet-based image restoration,” *IEEE Trans. Image Proc.*, vol. 12, no. 8, pp. 906–916, August 2003.
- [26] J Bioucas-Dias, “Bayesian wavelet-based image deconvolution: a gem algorithm exploiting a class of heavy-tailed priors,” *IEEE Trans. Image Proc.*, vol. 4, pp. 937–951, April 2006.
- [27] R Molina, J Mateos, A K Katsaggelos, and M Vega, “Bayesian multichannel image restoration using compound gauss-markov random fields,” *IEEE Trans. Image Proc.*, vol. 12, no. 12, pp. 1642–1654, December 2003.
- [28] M R Banham and A K Katsaggelos, “Spatially adaptive wavelet-based multiscale image restoration,” *IEEE Trans. Image Proc.*, vol. 5, pp. 619–634, April 1996.
- [29] R Neelamani, H Choi, and R G Baraniuk, “Wavelet-domain regularized deconvolution for ill-conditioned systems,” in *IEEE Int'l Conf. on Image Proc.*, Kobe, Japan, October 1999, vol. 1, pp. 204–208.
- [30] R Neelamani, H Choi, and R G Baraniuk, “ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems,” *IEEE Trans. Signal Proc.*, vol. 52, no. 2, pp. 418–433, February 2004.
- [31] J Kalifa and S Mallat, “Mini-max restoration and deconvolution,” in *Bayesian inference in wavelet based methods*. Springer, 1999.
- [32] A Jalobeanu, N Kingsbury, and J Zerubia, “Image deconvolution using hidden markov tree modeling of complex wavelet packets,” in *IEEE Int'l Conf. on Image Proc.*, 2001, vol. 1, pp. 201–204.
- [33] A Foi, K Dabov, V Katkovnik, and K Egiazarian, “Shape-adaptive dct for denoising and image reconstruction,” in *SPIE Electronic Imaging 2006, Image Processing: Algorithms and Systems V*, 2006, pp. 6064A–18.
- [34] A W Stevenson, T E Gureyev, Y I Nesterets and S W Wilkins, “A method for local deconvolution,” *Applied Optics*, vol. 42, no. 32, pp. 6488–6494, November 2003.
- [35] J Portilla and E P Simoncelli, “Image restoration using Gaussian scale mixtures in the wavelet domain,” in *IEEE Int'l Conf. on Image Proc.*, Barcelona, Spain, September 2003, vol. 2, pp. 965–968.
- [36] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger, “Shifttable multi-scale transforms,” *IEEE Trans Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992, Special Issue on Wavelets.
- [37] R R Coifman and D L Donoho, “Translation-invariant de-noising,” in *Wavelets and statistics*, A Antoniadis and G Oppenheim, Eds. Springer-Verlag lecture notes, San Diego, 1995.
- [38] E P Simoncelli and W T Freeman, “The steerable pyramid: A flexible architecture for multi-scale derivative computation,” in *IEEE Int'l Conf. on Image Proc.*, Washington, DC, October 1995, vol. III, pp. 444–447, IEEE Sig Proc Society.
- [39] E Candès and D Donoho, *Curvelets: A Surprisingly Effective Nonadaptive Representation of Objects with Edges*, Vanderbilt University Press, Nashville, TN, 1999.
- [40] N Kingsbury, “Image processing with complex wavelets,” September 1999, Phil. Trans. Royal Society London A.
- [41] J L Starck, M Elad, and D L Donoho, “Image decomposition via the combination of sparse representations and a variational approach,” *IEEE Trans. Image Proc.*, vol. 14, no. 10, pp. 1570–1582, 2005.
- [42] J L Starck, D L Donoho, and E Candes, “Very high quality image restoration,” in *Proc. SPIE conf. Signal and Image Processing.*, San Diego, August 2001, vol. 4478, pp. 9–19.
- [43] J K Romberg, M B Wakin, and R G Baraniuk, “Multiscale geometric image processing,” in *in Proc. of the SPIE*, 2005, vol. 5150, pp. 1265–1272.
- [44] E L Pennec and S Mallat, “Sparse geometric image representations with bandelets,” *IEEE Trans. Image Proc.*, vol. 14, no. 4, pp. 423–438, April 2005.
- [45] A Haar, “Zur theorie der orthogonalen funktionensysteme,” *Annals of Mathematics*, vol. 69, pp. 331–371, 1910.
- [46] E R Kretzmer, “The statistics of television signals,” *Bell System Tech. J.*, vol. 31, pp. 751–763, 1952.
- [47] N G Deriugin, “The power spectrum and the correlation function of the television signal,” *Telecommunications*, vol. 1, no. 7, pp. 1–12, 1956.
- [48] E P Simoncelli, “Statistical models for images: Compression, restoration and synthesis,” in *Proc 31st Asilomar Conf on Signals, Systems and Computers*, Pacific Grove, CA, November 1997, pp. 673–678.
- [49] J Shapiro, “Embedded image coding using zerotrees of wavelet coefficients,” *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3445–3462, December 1993.
- [50] R W Buccigrossi, *Compression and Segmentation of Images Using an Inter-Subband Wavelet Probability Model*, Ph.D. thesis, University of Pennsylvania, Department of Computer and Information Science, Philadelphia PA, June 1999.
- [51] J Portilla and E P Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–70, 2000.
- [52] S Lyu and E P Simoncelli, “Statistical modeling of images with fields of gaussian scale mixtures,” in *Adv. Neural Information Processing Systems*. May 2007, vol. 19.
- [53] E Vansteenkiste, D Van der Weken, W Philips, and E E Kerre, “Perceived image quality measurement of state-of-the-art noise reduction schemes,” in *Advanced Concepts for Intelligent Vision Systems*, 2006, pp. 114–126.
- [54] J Liu and P Moulin, “Complexity-regularized image restoration,” in *IEEE Int'l Conf. on Image Proc.*, 1998, vol. 1, pp. 555–559.
- [55] J Portilla, J A Guerrero-Colon, and C Dorronsoro, “Patente: Procedimiento de restauracion de imagenes afectadas por imperfecciones, dispositivo para llevarlo a cabo y sus aplicaciones,” 2006, Pending, Filled on August 2006, P2006021116.
- [56] M J Shensa, “The discrete wavelet transform: wedding the a trous and mallat algorithms,” *IEEE Trans. Signal Proc.*, vol. 40, no. 10, pp. 2464–2482, October 1992.



Jose A. Guerrero-Colón was born in Málaga, Spain. He received his MS Degree in Computer Science in 2003 from the Escuela Técnica Superior de Ingeniería Informática, Universidad de Málaga, Spain. Currently his carrying out his Ph.D. studies and working in several projects related to his main topic: New visual-statistical models applied to the restoration of photographic images corrupted by blur and noise. He is a grant holder of FPU contract, funded by Ministerio de Educación y Ciencia, from the Spanish government, at the Dept. of Computer

Science and Artificial Intelligence, Universidad de Granada, Spain.



Luis Mancera received his MS Degree in Computer Science in 2001 from the Escuela Técnica Superior de Ingeniería Informática in the Universidad de Granada, Spain. He worked for two years for the European Centre for Nuclear Research (CERN) in Geneva, Switzerland. Now he is carrying out his Ph.D. studies mainly centered in non-linear sparse image representation methods and their applications. He is a grant holder of a FPI contract at the Dept. of Computer Science and Artificial Intelligence, Universidad de Granada, Spain, funded by Ministerio de

Educación y Ciencia, from the Spanish government.



Javier Portilla graduated (M.S.'94, Ph.D.'99) from E.T.S.I. Telecomunicación, in the Universidad Politécnica de Madrid. After a post-doctoral stay in New York University, he worked as an Assistant Professor at the Computer Science and Artificial Intelligence Department, in the Universidad de Granada, Spain. He obtained a “Ramón y Cajal” research position, which he still enjoys, now in the Instituto de Óptica, Consejo Superior de Investigaciones Científicas, back in Madrid. He is interested on the one hand, on image statistical modelling and their degradation sources, for designing Bayesian methods, and, on the other hand, on new general purpose non-linear image representation schemes inspired in biological vision, and their application to image processing.