

TWO-LEVEL ADAPTIVE DENOISING USING GAUSSIAN SCALE MIXTURES IN OVERCOMPLETE ORIENTED PYRAMIDS

Jose A. Guerrero-Colon* and Javier Portilla†

Visual Information Processing Group
Dept. of Computer Science and Artificial Intelligence
Universidad de Granada
{jaguerrero, javier}@decsai.ugr.es

We describe an adaptive denoising method for images decomposed in overcomplete oriented pyramids. Our approach integrates two kinds of adaptation: 1) a 'coarse' adaptation, where a large window is used within each subband to estimate the local signal covariance; 2) a 'fine' adaptation, which uses small neighborhoods of coefficients modelled as the product of a Gaussian and a hidden multiplier, i.e., as Gaussian scale mixtures (GSM). The former provides adaptation to local spectral features, whereas the latter adapts to local energy fluctuations. We formulate our method as a Bayes Least Squares estimator using spatially variant GSMs. We also discuss the importance of image representation, compare the results using two different representations with complementary features, and study the effect of merging their results. We demonstrate through simulation that our method surpasses the state-of-the-art performance, in a L_2 -norm sense.

1. INTRODUCTION

Spatially adaptive models have been widely used for image denoising. In 1980, Lee proposed an adaptive Wiener method in the pixel domain [1]. With the introduction of *wavelets* (in the broad sense of the word), the local adaptation in the new domain became much more powerful. Some authors (e.g. [2, 3, 4]) set local thresholds for the coefficients depending on contextual measurements. Other authors continued developing the local Wiener idea in the new domain. Mihçak et. al. [5] included a prior on the local variance. A similar Bayesian approach was followed in [6]. Simoncelli et. al. [7] refined the Bayesian model using full signal and noise covariance, under a new Gaussian scale mixture (GSM) frame. All these methods followed an empirical Bayes strategy: first, estimate the local en-

ergy, and then, apply a Wiener filter using that estimation. Portilla et. al. [8], instead, introduced a single-step Bayes Least Squares estimator, further increasing the denoising performance. The referred adaptive Wiener methods either consider just the local variance, disregarding the covariance structure for each subband [5, 6], or consider the latter fixed, up to a scale factor [7, 8]. In both cases they perform a 'fine' adaptation, in the sense that local energy estimation is carried out using small neighborhoods.

Broad literature is available on local covariance estimation and its applications (e.g. [9]). In image denoising, this kind of estimation allows for a selective preservation of salient local spatial frequencies. However, as covariance estimation requires a much larger window than variance estimation, the result is a too 'coarse' adaptation. Thus, the extra complexity does not pay off in performance terms.

In this work we integrate both referred types of adaptation. We estimate the local covariance in large neighborhoods of the image subbands, and, once we are adapted to those features, we use small neighborhoods as in [8] to account for local variance fluctuations. This two-level adaptation results in a significant performance increase with respect to current state-of-the-art methods.

Besides adaptation, another critical issue is image representation. In this work we use two complementary representations: the full steerable pyramid [8], and an overcomplete version of the Haar wavelet [10]. We discuss the use of these representations, and, as in [11], show the potential benefits of combining their results.

2. IMAGE MODELLING

2.1. Image Representation

In recent years, the power of statistical image models has been increased by transforming the image from the pixel domain to new representations. Overcomplete oriented pyramids (such as the steerable pyramid, the curvelet or the complex wavelets), have demonstrated to better fit the statistical

*JAGC is supported by 2050-559 grant contract, funded by CIDA, Ministerio de Defensa (Spain).

†JP is supported by TIC2003-01504 grant, funded by the Ministerio de Ciencia y Tecnología (Spain) and the "Ramón y Cajal" program.

properties of typical images rather than single-scale representations. Each of the referred representations fits best for processing certain image features, but none is superior in absolute terms. For this reason, some authors have proposed to combine them. The simplest way to combine two representations is to average their results [11].

In this work we have used two representations: the Full Steerable Pyramid (FSP) and the Trapezoidal Undecimated Pyramid (TUP). The FSP is an extension of Simoncelli’s original steerable pyramid, including orientation selectivity in the high-pass residual [8]. The major advantages of the FSP are that (1) it is flexible in its number of orientations, and (2) their filters are highly selective in frequency. Its drawbacks are the large spatial support of the filters and the high computational cost derived from its large redundancy (best denoising results are reported for 8 orientations [8], a redundancy factor close to 19). TUP is an overcomplete version of the Haar wavelet, obtained by decimating its a *trous* subbands from the third scale on, counting upwards. This gives rise to a *trapezoidal* oriented pyramid with a redundancy factor of 7. Unlike the Haar wavelet, and like FSP, TUP is invariant to translation, and it still has (almost) perfect reconstruction (e.g., PSNR of 53.6 dB for *Lena* reconstruction). Complementary to FSP, TUP’s kernels are very small, thus providing excellent spatial localization. TUP’s weaknesses are its poor performance for processing oblique orientations and the reduced spectral localization of its kernels. According to these features, FSP is more powerful for processing texture-rich images, and TUP for processing images made of flat patches with well-defined boundaries (especially if oriented horizontally and vertically). Due to their complementary character, averaging the results obtained with each representations will generally increase the overall performance.

2.2. Spatially Variant GSM

The GSM model in the wavelet domain [12] is able to represent local variance fluctuations, a basic feature of typical images that explain most of the non-linear statistical patterns reported in recent years. This model assumes a fixed signal covariance for all neighborhoods in the same subband. However, typical images have different spatial covariance at different locations. Note also that, although the filtering associated with each subband reduces the spatial variability of the covariance within a subband, it does not remove it completely. Thus, to further increase the adaptability of our model, we use a GSM model with a signal covariance that changes at different regions of the image.

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 4). We model these neighborhoods as Gaussian scale mixtures (GSMs). A random vector \mathbf{x} is a GSM [13] 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 spatially variant behavior by defining $\mathbf{C}_u^{j,(n,m)}$ as the (spatially variant) covariance of $\mathbf{u}^j(n, m)$. The density of \mathbf{x} is determined uniquely by $p_z(z)$ and $\mathbf{C}_u^{j,(n,m)}$ as follows:

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

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 . Following [8], we have assumed the mixing density $p_z(z)$ constant in $\log z$ for all the subbands in the working range.

3. TWO-LEVEL ADAPTIVE DENOISING

We assume the image is corrupted by additive zero-mean independent Gaussian noise of known spectral density. Our image denoising method, as it is usual, performs the denoising in the *wavelet* domain and then invert the pyramid. Following the GSM model, we can express a neighborhood of noisy coefficients as $\mathbf{y} = \sqrt{z}\mathbf{u} + \mathbf{w}$ where \mathbf{w} is the zero mean Gaussian noise vector with covariance matrix \mathbf{C}_w . Note that \mathbf{C}_w (unlike the signal covariance \mathbf{C}_u) is assumed constant for all neighborhoods within a subband (but different for each subband).

3.1. Parameter estimation

For a given z , the covariance of a vector of observations is $\mathbf{C}_{y|z} = z\mathbf{C}_u + \mathbf{C}_w$. Because of the independence of z , \mathbf{u} and \mathbf{w} , and taking expectations over z , it yields: $\mathbf{C}_y = \mathbb{E}\{\mathbf{C}_{y|z}\} = \mathbb{E}\{z\}\mathbf{C}_u + \mathbf{C}_w$. Setting $\mathbb{E}\{z\} = 1$, we obtain $\mathbf{C}_y = \mathbf{C}_u + \mathbf{C}_w$. \mathbf{C}_w is computed at each subband by applying the pyramidal representation to a deterministic function with the same sample autocorrelation as the noise. To estimate the spatially variant signal covariance \mathbf{C}_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 regions. We divide each subband into $B \times B$ blocks, which are assumed uniform in covariance. 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 .

¹For notational simplicity, we drop the superscript j and the indices (n, m) in the following.

3.2. 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 [8] as

$$\mathbb{E}\{x_c|\mathbf{y}\} = \int_0^\infty p(z|\mathbf{y}) [\mathbb{E}\{\mathbf{x}|\mathbf{y}, z\}]_{x_c} dz. \quad (1)$$

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 (1) 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 [8]). Note that, unlike in [8] 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 for each block of each subband.

4. IMPLEMENTATION

We have used for both representations 4 scales for 256×256 images and 5 scales for 512×512 images. The number of orientations is fixed for the TUP (3), and 8 for the FSP as in [8]. For both representation we have used a 3×3 GSM neighborhood around each coefficient. In the case of FSP we also included a coefficient at the same location and orientation at the next coarser scale (a *parent*), as in [8]. We have tested different size blocks B for estimating the signal covariance ($B = \{16, 32, 64, 128\}$). We have used the same block size B for all pyramid levels. Finally, we have been very careful to avoid boundary artifacts, a critical issue for block-based image processing. We have used non-overlapping blocks, with overlapping extensions, which become mirror reflections at the image boundaries.

5. RESULTS AND DISCUSSION

In Table 1 we show the results of the method applied to four standard test images and four noise levels. We have included a comparison to the original BLS-GSM method (a single level of adaptation, *1LA*), both with the FSP and the TUP. These results show the performance gain derived from the two-level adaptation (*2LA*) strategy. It is significant, though, how some images experience a big improvement (e.g., *Barbara*, an image having widely varying high frequency texture), whereas in other cases the improvement

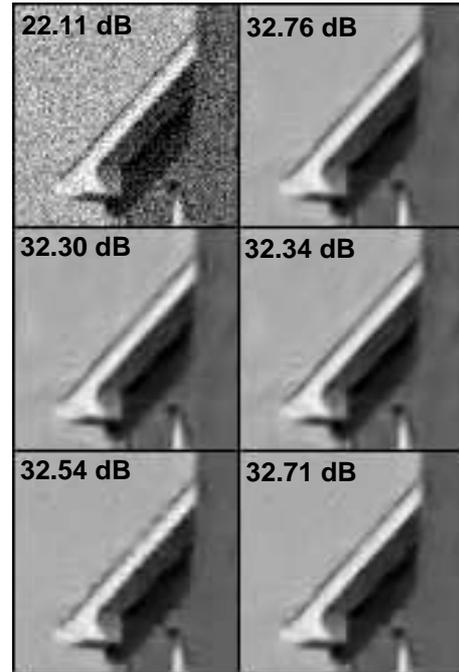


Fig. 1. Visual comparison results on *House* image (cropped to 80×80). From left to right and top to bottom: Noisy ($\sigma_w = 20$); Our FSP and TUP results averaged; BLS-GSM [8] with FSP; Our method with FSP; BLS-GSM [8] with TUP; Our method with TUP.

is modest (e.g., *Lena*, which does not present local spectral peaks). They also show how different representations may fit better different images, depending on their texture and edges contents: *House* and *Peppers* are best denoised using the TUP, whereas for *Lena* and *Barbara* we obtain better results with the FSP. The second column counting from the right shows the effect of averaging the denoised images obtained using each representation. We observe that, except for *Barbara*, this provides an extra performance gain, which in some cases (*House*) becomes very significant. Finally, the last column shows the block size used for the reported results (excluding the averaged ones), the one providing the best results for each case. We observe that this parameter depends on the particular image, and that it tends to increase with the noise level.

Fig. 1 shows some visual results (a 80×80 pixel patch of *House*). This figure allows to visually interpret: 1) the effect of our two-level adaptation (right column) vs. one-level adaptation (BLS-GSM, left column); 2) the effect of using different representations (FSP in the middle row, and TUP in the bottom row); and 3) the effect of averaging both of them (up right corner). Note how many artifacts on the roof line obtained with the TUP representation disappear when using two-level adaptation. Finally, Figure 2 shows a PSNR comparison between our results (from Table 1, excluding the averaged ones) and the previously best reported results [8, 14, 3, 4, 6].

σ / PSNR	FSP		TUP		FSP +	Block Size
	1LA	2LA	1LA	2LA	TUP	
<i>Lena</i>						
10/ 28.13	35.60	35.66	35.23	35.38	35.71	64
15/ 24.61	33.90	33.96	33.50	33.65	34.01	64
20/ 22.11	32.67	32.71	32.25	32.40	32.77	64
25/ 20.17	31.69	31.72	31.26	31.40	31.78	64
<i>Barbara</i>						
10/ 28.13	34.02	34.60	33.13	34.13	34.58	16
15/ 24.61	31.83	32.49	30.76	31.85	32.40	16
20/ 22.11	30.27	30.94	29.08	30.19	30.81	16
25/ 20.17	29.07	29.76	27.81	28.81	29.52	32
<i>House</i>						
10/ 28.13	35.29	35.44	35.32	35.56	35.72	16
15/ 24.61	33.54	33.64	33.73	33.94	34.01	32
20/ 22.11	32.30	32.34	32.54	32.71	32.76	32
25/ 20.17	31.32	31.30	31.59	31.71	31.76	32
<i>Peppers</i>						
10/ 28.13	33.74	33.81	34.06	34.24	34.28	16
15/ 24.61	31.7	31.72	32.02	32.18	32.19	16
20/ 22.11	30.27	30.24	30.57	30.67	30.69	16
25/ 20.17	29.17	29.10	29.45	29.50	29.54	16

Table 1. Denoising performance expressed as Peak Signal-to-Noise Ratio, $10 \log_{10}(255^2/\sigma_e^2)$ in dB, where σ_e^2 is the mean square error. First column shows the noise standard deviation and the corresponding PSNR. Next four columns correspond to FSP and TUP representations in the one-level adaptive (1LA) and two-level adaptive (2LA) version (highlighted the best of those 4 results). Next column shows the PSNR obtained using the average of the two 2LA results (highlighted if they are better than non-averaged results). Last column shows the block size used for the 2LA method.

6. REFERENCES

- [1] 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.
- [2] S G Chang, B Yu, and M Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," in *Fifth IEEE Int'l Conf on Image Proc.*, Chicago, October 1998, vol. 1, pp. 535–539, IEEE Computer Society.
- [3] A Pižurica, W Philips, I Lemahieu, and M Acheroy, "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.
- [4] L Şendur and I W Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Proc.*, vol. 50, no. 11, pp. 2744–2756, November 2002.
- [5] M K Mihçak, I Kozintsev, K Ramchandran, and P Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Trans. on Signal Processing*, vol. 6, no. 12, pp. 300–303, December 1999.

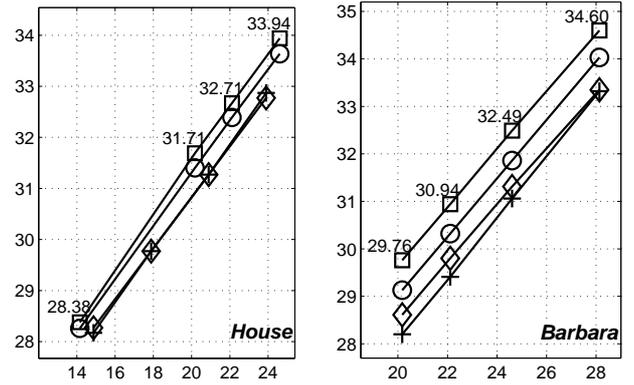


Fig. 2. PSNR output vs. PSNR input, in dB. We use TUP for *House* and FSP for *Barbara*. Our results, for $\sigma_w = \{10, 15, 20, 25\}$, are in Table 1. We use $B = 128$ for *House* with $\sigma_w = 50$. We compare to three state-of-the-art methods. *House*: [14] (crosses) and [3] (diamonds). *Barbara*: [6] (crosses) and [4] (diamonds). For both images: [8] (circles) and this method (squares).

- [6] 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, IEEE.
- [7] J Portilla, V Strela, M Wainwright, and E Simoncelli, "Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain," in *Proc 8th IEEE Int'l Conf on Image Proc.*, Thessaloniki, Greece, Oct 7-10 2001, pp. 37–40, IEEE Computer Society.
- [8] J Portilla, V Strela, M 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.
- [9] S. Mallat, G. Papanicolaou, and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *Ann. Stat.*, vol. 26, no. 1, pp. 1–47, 1998.
- [10] J. Portilla, "Full blind denoising through noise covariance estimation using gaussian scale mixtures in the wavelet domain," in *IEEE International Conference on Image Processing*, IEEE, Ed. Singapore, October 2004, pp. 1217–1220.
- [11] 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, A Laine, M Unser and A Aldroubi, Eds.
- [12] M J Wainwright and E P Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Adv. Neural Information Processing Systems*, S. A. Solla, T. K. Leen, and K.-R. Müller, Eds., Cambridge, MA, May 2000, vol. 12, pp. 855–861, MIT Press, Presented at Neural Information Processing Systems, Dec 1999.
- [13] D Andrews and C Mallows, "Scale mixtures of normal distributions," *J. Royal Stat. Soc.*, vol. 36, pp. 99–102, 1974.
- [14] M Malfait and D Roose, "Wavelet-based image denoising using a Markov random field a priori model," *IEEE Trans. Image Proc.*, vol. 6, pp. 549–565, April 1997.