BLIND NON-WHITE NOISE REMOVAL IN IMAGES USING GAUSSIAN SCALE MIXTURES IN THE WAVELET DOMAIN

Javier Portilla

Visual Information Processing Group
Dept. of Computer Science and A.I., Universidad de Granada, Spain
javier@decsai.ugr.es

ABSTRACT

Gaussian scale mixtures (GSM) capture two basic properties of the wavelet coefficients responding to natural images, namely 1) high kurtosis marginals, and 2) positive covariance between neighbor coefficient amplitudes. These features are not shared by Gaussian or lower kurtosis noise sources. Therefore, GSM models provide a means to separate the noise from the signal for an observed corrupted image. A local model consisting of a GSM term plus Gaussian additive noise with arbitrary covariance is used to estimate first the noise covariance at each wavelet subband, applying a generalized expectation maximization algorithm. Then the original wavelet coefficients are estimated from the noisy observations using an efficient Bayes Least Squares technique. Both steps are fully automatic.

1. INTRODUCTION

Blind noise removal usually requires the estimation of the noise statistics from the observation. Traditionally, noise is assumed white, which allows to rely on the different spectral behavior of white noise and natural images to estimate the noise variance. However, most often, noise in real degraded images is far from white. Note that if we remove the spectral constraints on both noise and signal we need to use higher-order statistics to discriminate between them (e.g., it would be impossible to distinguish between noise and signal if both were modelled as Gaussian of unknown covariance). Thus, any practical general purpose blind denoising method requires: 1) a noise model allowing non-flat spectral behavior; and 2) a signal model involving higher-order statistics. In this work we have used a correlated Gaussian model for the noise, and correlated Gaussian Scale Mixtures (locally, in an overcomplete wavelet domain) for the signal, as in [1]. However, unlike the referred work, we do not assume prior knowledge of the noise covariance, but apply instead a generalized expectation-maximization algorithm for estimating it from the degraded image.

This work is supported by the Ministerio de Ciencia y Tecnologa (MCYT, Spain) through grant TIC2003-01504.

2. MODELLING THE IMAGE STATISTICS

Multi-scale representation schemes have been extensively used during the last years for analyzing and processing natural signals (images, sounds, seismic waves, etc.). For signal restoration, in particular, it has been observed that redundant representations are more effective than critically sampled wavelets [2, 3]. For this work we have used an overcomplete version of the Haar wavelet that, similarly to the *à trous* representation is aliasing free, but, unlike the latter, it preserves the pyramid structure. Although this is the simplest option, any other aliasing-free oriented pyramid is also applicable (e.g., Simoncelli's steerable pyramid [1], or the curvelets [4]). Once we have decomposed the original image into oriented subbands at different scales, we observe two outstanding statistical features of the signal coefficients: (1) marginal statistics are strongly leptokurtotic (peaked at zero, with heavy tails) [5]; and (2) the amplitude of neighbor coefficients present a strong positive covariance [13].

2.1. Gaussian Scale Mixtures in the wavelet domain

We use Gaussian Scale Mixtures [6] (GSM) for modelling neighborhoods of coefficients for each subband [7]. In this work we have used neighborhoods of 3×3 coefficients. Every localized feature in the image domain gives raise to simultaneous responses of several coefficients in the wavelet domain. The underlying idea of the GSM model is that every small cluster of coefficients can be seen as a Gaussian vector sample scaled in amplitude by an unknown factor that modulates the local energy: $\mathbf{x} = \sqrt{z}\mathbf{u}$, where $\mathbf{x}$ represents the neighborhood of image coefficients for the subband, $\sqrt{z}$ is the unknown positive scalar and $\mathbf{u}$ is a Gaussian vector, independent of $\sqrt{z}$. This simple model reproduces the two basic features explained above. First, providing that the density of z is concentrated around 0, with a heavy positive tail, most often $\mathbf{x}$ will have a small magnitude (soft texture areas), but it will occasionally produce high magnitude responses (edges, lines, corners, etc.). Second, when we observe a

coefficient having a small amplitude, it will most likely be because z is small, and then its neighbors will most likely be small as well. Similarly, when we observe a large amplitude coefficient, it most likely indicates a large z (i.e., a high local-contrast feature), which means that its neighbors will have a large variance. This is exactly what has been reported to happen in real images [13].

3. NOISE COVARIANCE ESTIMATION

3.1. Image degradation model

For the degradation, we consider stationary zero-mean independent additive Gaussian noise of arbitrary autocovariance $\mathbf{a_w}$. This produces zero-mean additive Gaussian noise ($\mathbf{w}$) in each subband of the pyramid, but with a different noise covariance matrix $\mathbf{C_w}$ for each subband. The vector of a neighborhood of noisy coefficients is $\mathbf{y} = \mathbf{x} + \mathbf{w} = \sqrt{z}\mathbf{u} + \mathbf{w}$. Being $\mathbf{u}$ and $\mathbf{w}$ Gaussian and independent, $p_{\mathbf{y}|z}(\mathbf{y}|z)$ is also Gaussian and the observation $\mathbf{y}$, as $\mathbf{x}$, is an infinite mixture of Gaussian vectors:

$$p_{\mathbf{y}}(\mathbf{y}) = \int_0^\infty p_{\mathbf{y}|z}(\mathbf{y}|z)\, p_z(z)\, dz. \tag{1}$$

Without loss of generality we set $\mathbb{E}\{z\} = 1$, which implies $\mathbf{C_x} = \mathbb{E}\{z\}\mathbf{C_u} = \mathbf{C_u}$. Note also that $\mathbf{C_x} = \mathbf{C_y} - \mathbf{C_w}$. Then the covariance of $\mathbf{y}$ for a given z is $\mathbf{C}_{\mathbf{y}|z} = z\mathbf{C_u} + \mathbf{C_w} = z\mathbf{C_y} + (1-z)\mathbf{C_w}$, and we can write

$$p_{\mathbf{y}|z}(\mathbf{y}|z) = \frac{\exp\left(-\mathbf{y}^T(z\mathbf{C_y} + (1-z)\mathbf{C_w})^{-1}\mathbf{y}/2\right)}{(2\pi)^{N/2}|z\mathbf{C_y} + (1-z)\mathbf{C_w}|^{1/2}} \tag{2}$$

where N is the number of coefficients in the neighborhood (9 in our case). The unknown features of $p_{\mathbf{y}}(\mathbf{y})$ (density of the observed vectors for each subband) are the noise covariance matrix $\mathbf{C_w}$ and the mixing density $p_z(z)$.

3.2. Estimating the model parameters

From an initial guess, we look for an updating rule such that the likelihood of the observations according to the model increases. We use the $Q(\mathbf{C_w}, \mathbf{C_w^{old}})$ criterion [12]

$$Q = \sum_{m=1}^{M} \int_0^\infty p_{z|\mathbf{y}}(z|\mathbf{y}_m; \mathbf{C_w^{old}}) \log(p(\mathbf{y}_m, z; \mathbf{C_w}))dz.$$

Using $\log(p(\mathbf{y}_m, z)) = \log(p_{\mathbf{y}|z}(\mathbf{y}_m|z)) + \log(p_z(z))$ and noting that $p_z(z)$ does not depend on $\mathbf{C_w}$, we express the gradient of $Q(\mathbf{C_w}, \mathbf{C_w^{old}})$ w.r.t. $\mathbf{C_w}$ as:

$$\frac{\partial Q}{\partial \mathbf{C_w}} = \sum_{m=1}^{M} \int_0^\infty p_{z|\mathbf{y}}(z|\mathbf{y}_m; \mathbf{C_w^{old}}) \frac{\partial \log(p_{\mathbf{y}|z}(\mathbf{y}_m|z; \mathbf{C_w}))}{\partial \mathbf{C_w}} dz.$$

After operating in the previous equation [9] we obtain:

$$\frac{\partial Q}{\partial \mathbf{C_w}} = \frac{M}{2} \int_0^\infty p_z^{new}(z)(1-z)\mathbf{C}_z^{-1}(\mathbf{I} - \widehat{\mathbf{C}_z}\mathbf{C}_z^{-1})dz, \tag{3}$$

where

$$p_z^{new}(z) = \frac{1}{M}\sum_{m=1}^{M} p_{z|\mathbf{y}}(z|\mathbf{y}_m; p_z^{old}(z), \mathbf{C_w^{old}}) \tag{4}$$

$$\mathbf{C}_z = \mathbf{C}_{\mathbf{y}|z} = z\mathbf{C_y} + (1-z)\mathbf{C_w} \tag{5}$$

$$\widehat{\mathbf{C}_z} = \frac{\sum_{m=1}^{M} p_{z|\mathbf{y}}(z|\mathbf{y}_m; p_z^{old}(z), \mathbf{C_w^{old}})\mathbf{y}_m\mathbf{y}_m^T}{\sum_{m=1}^{M} p_{z|\mathbf{y}}(z|\mathbf{y}_m; p_z^{old}(z), \mathbf{C_w^{old}})}.$$

$\mathbf{I}$ is the N-identity matrix and $p_z^{new}(z)$ is estimated through Eq. 4. The latter equation is formally identical to the EM solution for Gaussian mixtures (e.g., [11]). Unfortunately, equating the r.h.s. of Eq. 3 to $\mathbf{0}$ and solving for $\mathbf{C_w}$, as the maximization step of an EM method, is challenging. Instead we just search for a $\mathbf{C_w^{new}}$ accomplishing $Q(\mathbf{C_w^{new}}, \mathbf{C_w^{old}}) > Q(\mathbf{C_w^{old}}, \mathbf{C_w^{old}})$. A possibility for updating the estimate is following the gradient direction:

$$\mathbf{C_w^{new}} = \mathbf{C_w^{old}} + \eta \frac{\partial Q(\mathbf{C_w}, \mathbf{C_w^{old}})}{\partial \mathbf{C_w}} \mid_{\mathbf{C_w}=\mathbf{C_w^{old}}}, \tag{6}$$

where η represents a suitable scale factor. However, as the evaluation of $Q(\mathbf{C_w}, \mathbf{C_w^{old}})$ is costly, gradient ascent is not an efficient solution. Instead, we observe: 1) $\mathbf{C}_z|_{z=0} = \mathbf{C_w}$; and 2) $\widehat{\mathbf{C}_z}$ could be regarded as ML-estimates of the covariance matrices $\mathbf{C}_z$ (Gaussian mixture EM solutions [11]), if only those matrices were not coupled to each other through Eq. 5. Thus, $\widehat{\mathbf{C}_z}|_{z=0}$ is a reasonable suboptimal estimate for $\mathbf{C_w}$. Then, we choose the updating rule:

$$\mathbf{C_w^{new}} = \frac{\sum_{m=1}^{M} p(0|\mathbf{y}_m; p_z^{old}(z), \mathbf{C_w^{old}})\mathbf{y}_m\mathbf{y}_m^T}{\sum_{m=1}^{M} p(0|\mathbf{y}_m; p_z^{old}(z), \mathbf{C_w^{old}})}. \tag{7}$$

Note that the posteriors $p(0|\mathbf{y}_m)$ will only be significant at the image locations where the noise clearly dominates over the signal. We have found that Eq. 7, together with Eq. 4, provides a very fast fitting of the model to the data. However, Eq. 7 does not guarantee a monotonic increase in Q, so it must be replaced by Eq. 6 for the rare cases where it diminishes Q.

4. FULL BLIND DENOISING

For every wavelet subband, once we have estimated the noise covariance and the mixing density, we apply the BLS-GSM method [1] for cleaning it (BLS stands for Bayes Least Squares). The final result is obtained by reconstructing an image from the processed subbands.

We remind the reader the two basic assumptions made for the noise: (1) it is additive, independent of the signal and stationary (spatially homogeneous); and (2) it is Gaussian. The second constraint can be relaxed in practice, because the signal model can only represent the leptokurtotic (kurtosis > 3) component of the observation, and thus, any remaining mesokurtotic (kurtosis ≤ 3) component is identified as noise by the model.

5. RESULTS AND DISCUSSION

In [8] we use simulations to demonstrate the performance of this method. Here we focus on applying the algorithm to real degraded images. We include noisy pictures captured using four different imaging devices. They, together with their denoised counterparts, are shown in Fig.1. From top to bottom, we label them as **(a)** to **(d)**[1]. **(a)** is the result of scanning a printed image, presenting the typical Moiré artifacts. It was obtained from the web site of NeatImage©[2]. We see how the method performs well for this regular and homogeneous degradation. **(b)** is an infrared video frame captured with an FPA IR video camera. The main source of degradation is an interference pattern, caused by a synchronism problem. This is an extreme case, where the noise spectral density approaches to a pair of deltas in Fourier. Nonetheless, the model is still applicable, and the result is very satisfactory. **(c)** is a digital photograph (Canon Powershot A70, 400 ISO) taken at night, with poor artificial illumination. Digital photographs are challenging for noise removal, because their noise is strongly dependent on the signal [10], which breaks our noise homogeneity assumption. Thus, for optimal results, a calibration of the noise generation features of the camera is required [10]. Furthermore, the pictures present occasional isolated *stuck* pixels, which also breaks the mesokurtotic assumption for the noise. Still, we obtain a strong noise reduction. Note the apparent decrease in contrast of the denoised image. This is due to more low frequency components of the signal becoming compatible with the mesokurtotic noise constraint as we go up in scale to remove the low frequency noise. Thus, some low frequencies of the signal are damped as well. **(d)** is an Optical Coherence Tomography of a pig's cornea. It is a challenging example because of the spatial inhomogeneity of the noise (smaller grains near the top). This provokes that the cleaned image presents some remaining artifacts at the areas where the local noise statistics differ from the dominant ones.

[1] **(a)** and **(c)** were originally in color, whereas **(b)** was colored with a palette for helping the termal interpretation (see the images in color at the electronic version of these proceedings).

[2] That web site shows some spectacular denoising results. Note that, unlike our denoising algorithm, their method requires human intervention and/or previous calibration of the imaging devices.

6. ACKNOWLEDGEMENTS

We are thankful to Carlos Dorronsoro and collaborators (CIDA, Defense Ministry, Spain) for providing us infrared video frames; Susana Marcos and collaborators (Visual Optics and Biophotonics Laboratory, CSIC, Spain) for providing us the OCT image; Filip Rooms and Wilfried Philips (TELIN, Ghent University, Belgium) for providing us noisy digital photographs and for the invitation to this symposium, respectively; and NeatImage© for making available to the public original and processed degraded images. We also thank Eero Simoncelli and Rafael Molina for useful discussions.

7. REFERENCES

[1] 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, no. 11, pp. 1338–1351, November 2003.

[2] E P Simoncelli, W T Freeman, E H Adelson, and D J Heeger, "Shiftable multi-scale transforms," *IEEE Trans Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992, Special Issue on Wavelets.

[3] R R Coifman and D L Donoho, "Translation–invariant de–noising," in *Wavelets and statistics*, Springer-Verlag lecture notes, San Diego, 1995.

[4] J Starck, E J Candes, and D L Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Proc.*, vol. 11, no. 6, pp. 670–684, June 2002.

[5] D J Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[6] D Andrews and C Mallows, "Scale mixtures of normal distributions," *J. Royal Stat. Soc.*, vol. 36, pp. 99–, 1974.

[7] M J Wainwright and E P Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Adv. Neural Information Processing Systems*, May 2000, vol. 12, pp. 855–861, MIT Press,

[8] J Portilla, "Full blind denoising through noise covariance estimation using Gaussian scale mixtures in the wavelet domain," Submitted to IEEE Int. Conf. on Im. Proc., 2004.

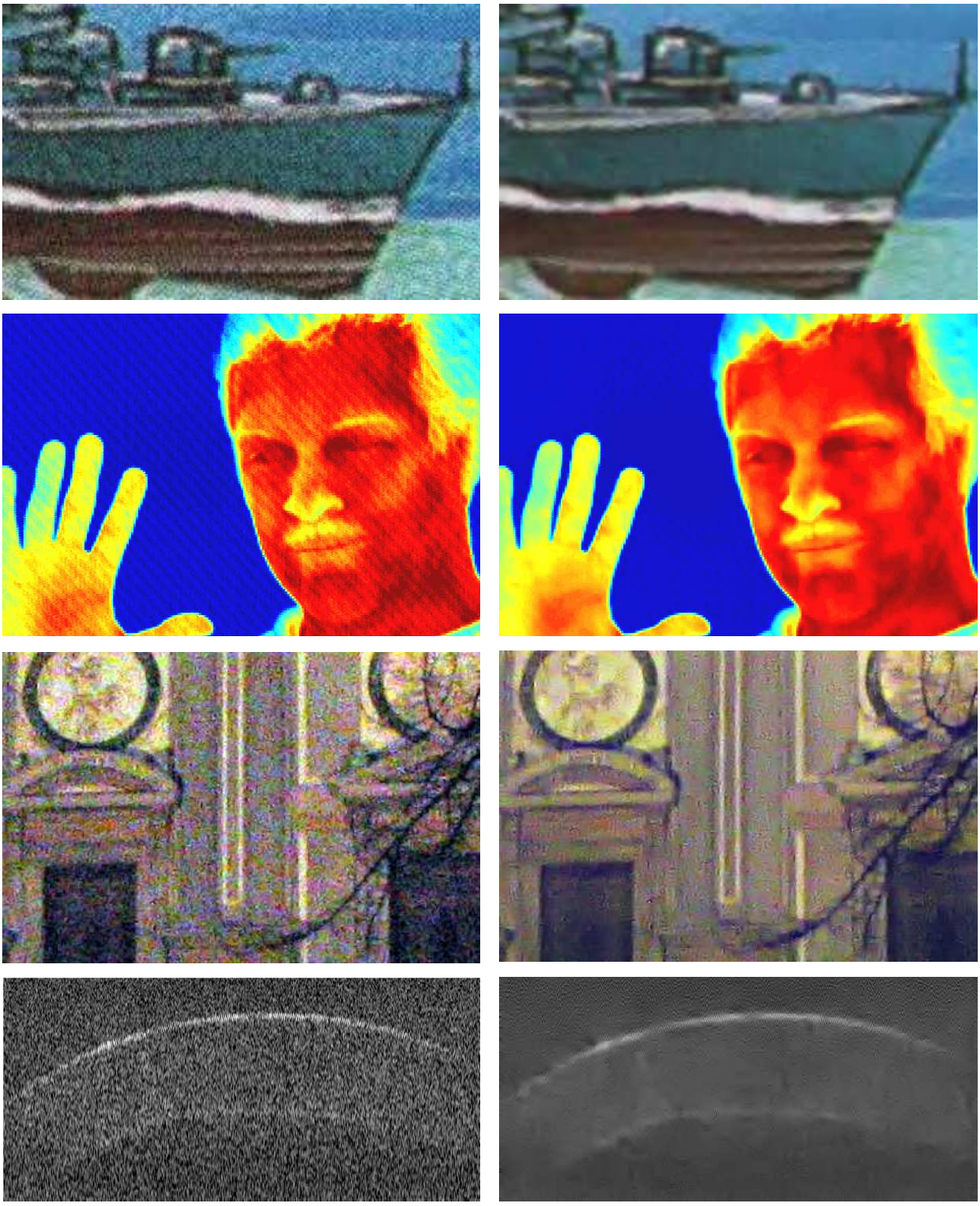
[9] J Portilla, "Full blind Bayesian denoising using scale mixtures of Gaussians in the wavelet domain," Tech. Rep. in preparation, DECSAI, Universidad de Granada, 2004.

[10] J Portilla, V Strela, M Wainwright, and E P Simoncelli, "Image denoising using Gaussian scale mixtures in the wavelet domain," Tech. Rep. TR2002-831, Courant Institute of Mathematical Sciences, N.Y.U., 2002.

[11] R O Duda, P E Hart, and D G Stork, *Pattern Classification, 2nd Edition*, chapter Unsupervised Learning and Clustering, Wiley Interscience, 2001.

[12] G J McLachlan and T Krishnan, *The EM algorithm and extensions*, Wiley, New York, 1996.

[13] E P Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *31st Asilomar Conf on Signals, Systems and Computers*, Pacific Grove, CA, November 1997, pp. 673–678



Observed *After Blind BLS-GSM Denoising*

Figure 1. Results obtained with four real noisy images. From top to bottom: **(a)** a printed image captured with a scanner; **(b)** an infrared video frame suffering from electronic interference; **(c)** a digital photography captured under poor light conditions; **(d)** an optical coherence tomography (OCT). Images have been zoomed and/or cropped for visibility of the artifacts. See text for more information.