

## IMAGE DENOISING USING MIXTURES OF GAUSSIAN SCALE MIXTURES

Jose A. Guerrero-Colón

Dept. of Comp. Science and A.I.  
Universidad de Granada  
Granada (Spain)

jaguerrero@decsai.ugr.es

Eero P. Simoncelli

Center for Neural Science, and  
Courant Inst. of Math. Sciences  
New York University (USA)

eero.simoncelli@nyu.edu

Javier Portilla

Instituto de Óptica  
CSIC  
Madrid (Spain)

portilla@io.cfmac.csic.es

*The local statistical properties of photographic images, when represented in a multi-scale basis, have been described using Gaussian scale mixtures (GSMs). In that model, each spatial neighborhood of coefficients is described as a Gaussian random vector modulated by a random hidden positive scaling variable. Here, we introduce a more powerful model in which neighborhoods of each subband are described as a finite mixture of GSMs. We develop methods to learn the mixing densities and covariance matrices associated with each of the GSM components from a single image, and show that this process naturally segments the image into regions of similar content. The model parameters can also be learned in the presence of additive Gaussian noise, and the resulting fitted model may be used as a prior for Bayesian noise removal. Simulations demonstrate this model substantially outperforms the original GSM model.*

**Index Terms**— Image denoising, Image modelling, Gaussian scale mixture.

### 1. INTRODUCTION

Natural images are highly inhomogeneous: smooth regions are interrupted by features such as texture, edges, lines, corners, etc. This heterogeneity produces highly non-Gaussian statistics, which are readily apparent when the image is transformed from the pixel domain to a linear multi-scale representation. Marginal histograms of coefficients reveal highly kurtotic behavior with heavy tails [1, 2]. In addition, there is a strong non-linear coupling between amplitudes of coefficients at nearby positions, orientations, and scales [3, 4]. A Gaussian scale mixture (GSM) model [5], in which clusters of coefficients are modeled as a product of a Gaussian vector and a positive scaling variable, captures both the amplitude coupling and the kurtotic marginals of wavelet coefficients [6]. Since the subbands are computed through band-pass filtering, it seems reasonable to assume that spatial neighborhoods within a subband share the same local covariance [7]. Better modeling may be obtained by allowing the neighborhoods of the image subbands to have a spatially varying covariance matrix. In previous work, this has been achieved by estimating covariance over a subregion of the image surrounding each neighborhood (known as SVGSM) [8], or by defining the covariances in a local coordinate system rotated to match the local dominant orientation (known as OAGSM) [9].

Several recent denoising methods demonstrate impressive results by exploiting the similarity of neighborhood structures scat-

tered throughout a given image [10, 11, 12]. Here, we introduce a finite mixture of GSM (MGSM) model that can effectively adapt the local covariance in a GSM model to nonlocal image structures. We develop an algorithm for adapting the parameters (in particular, the covariances of each of the components) to the data in a single image. We show first that fitting the model to an image produces a natural segmentation into regions of similar content. We then develop a Bayesian Least Squares estimator for removing additive Gaussian noise. We demonstrate through simulations that the resulting denoising performance is substantially better than that of a single GSM model (as reported in [7]), and nearly as good as the best results reported in recent literature.

### 2. MODELLING NON-LOCAL STATISTICS: MGSM

The modeling capabilities of a GSM can be significantly improved by estimating the covariance matrices locally [8, 9], because the neighborhoods over which the estimation is performed are more likely to be statistically similar. But this reduces the number of neighborhoods used in the estimate (thus increasing the estimation error), and still does not *guarantee* that they are not contaminated by neighborhoods drawn from different statistical sources (e.g., when estimated near a texture boundary). Instead, we would like to identify the dominant statistical structures occurring throughout the image, and then compute the contribution of each of these identified "modes" to the contents of each local neighborhood. We achieve this by introducing a *mixture of Gaussian scale mixtures* (MGSM) model. As with the GSM model, we write a (vectorized) coefficient neighborhood as:

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

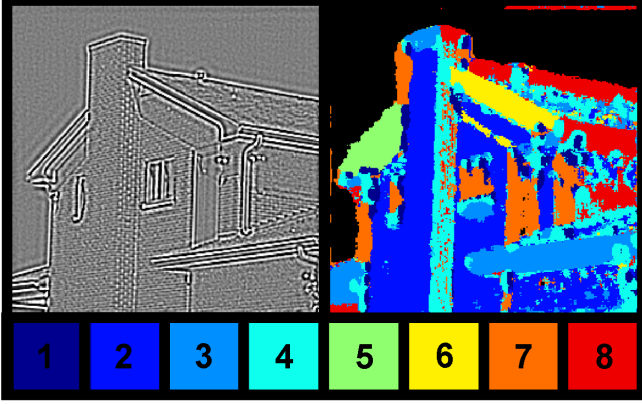
but unlike a simple GSM, we assume a hidden discrete index variable  $k \sim P_k$  is selected for each neighborhood and determines the mixing density of the scalar multiplier  $p(z|k) = p_k(z)$ , as well as the covariance matrix  $\mathbf{C}_k$  of zero-mean Gaussian vector  $\mathbf{u}$  given  $k$ . We assume  $\mathbf{u}$  and  $z$  are independent when conditioned on  $k$ . According to this, the density of the modelled neighborhood vector is:

$$\begin{aligned} p(\mathbf{x}) &= \sum_{k=1}^K P_k p(\mathbf{x}|k) = \sum_{k=1}^K P_k \int_z p(\mathbf{x}|k, z) p_k(z) dz \\ &= \sum_{k=1}^K P_k \int_z \frac{\exp(-\mathbf{x}^T(z\mathbf{C}_k)^{-1}\mathbf{x}/2)}{(2\pi)^{N/2}|z\mathbf{C}_k|^{1/2}} p_k(z) dz. \end{aligned} \quad (1)$$

In this paper, we model neighborhoods for each subband separately. For each subband, and assuming a neighborhood size  $N$  and number of mixture components  $K$ , the parameters of the MGSM model are: (a) the  $K$  probability masses  $P_k$ ; (b) the  $K$  mixing scale densities

---

JAGC was supported by AP2005-2969 FPU contract and TEC2006/13845/TCM grant, both funded by the Ministerio de Educación y Ciencia (Spain). EPS was supported by the Howard Hughes Medical institute. JP was supported by TEC2006/13845/TCM grant and the Ramón y Cajal Program, both funded by the Ministerio de Educación y Ciencia (Spain).



**Fig. 1.** Image subband labelled according to dominant MGSM component. Left: a non-oriented steerable pyramid subband of the *House* image. Right: color-coded indication of dominant component in an MGSM model with 8 components and  $7 \times 7$  neighborhoods. Black pixels indicate neighborhoods with energy below a threshold.

$p_k(z)$ ; and (c) the  $K$  covariance matrices  $\mathbf{C}_k$ . For a given image, we estimate these parameters by maximizing the likelihood of the data given the model. The procedure is an iterative EM-like algorithm, and a very similar version of it is described in Section 3 for the case when there is additive Gaussian noise in the observed image.

The MGSM model is able to capture similarities in neighborhood structures, even if they are non-adjacent. This is illustrated in Fig. 1, which shows a non-oriented image subband together with the result of fitting the MGSM model with  $K = 8$  and a  $7 \times 7$  GSM neighborhood. The color of each pixel identifies the dominant GSM mixture component for the neighborhood centered at that position. Specifically, we estimated the model parameters and then chose, for each neighborhood, the index  $k$  maximizing the posterior  $P(k|\mathbf{x})$ . Some mixture components can be seen to represent edges of a given orientation (components 3, 5, 6, 7 and 8), some represent homogeneous texture regions (2 and 4), and one captures corners (1). We find it remarkable that this model, which does not include any explicit definition of features or attributes, can perform an automatic segmentation of the image contents into these feature classes.

### 3. IMAGE DENOISING USING MGSM

The MGSM model provides a useful substrate for the application of denoising, where it can be used as a prior model in a Bayesian estimation scheme. Consider an image corrupted by additive Gaussian noise of known covariance (not necessary white). As with many other methods, we estimate the original image by transforming the noisy observation into an overcomplete multi-scale representation, denoising each subband, and then inverting the transform. We model the noisy neighborhoods of each subband as a mixture of Gaussian scale mixtures, plus a Gaussian noise term:

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

where  $\mathbf{w}$  is a zero mean Gaussian noise vector with covariance matrix  $\mathbf{C}_w$ . The density of the observed neighborhood vector conditioned on  $z$  and  $k$  is Gaussian (zero-mean), with covariance  $\mathbf{C}_{\mathbf{y}|z,k} = z\mathbf{C}_k + \mathbf{C}_w$ . Hence, the density of the observed vector  $\mathbf{y}$

is:

$$p(\mathbf{y}) = \sum_{k=1}^K P_k \int_z \frac{\exp(-\mathbf{y}^T(z\mathbf{C}_k + \mathbf{C}_w)^{-1}\mathbf{y}/2)}{(2\pi)^{N/2}|z\mathbf{C}_k + \mathbf{C}_w|^{1/2}} p_k(z) dz.$$

#### 3.1. BLS Coefficient estimation

As in previous work, we compute the Bayesian Least Squares (BLS) estimate (also known as the MMSE) of each coefficient given the surrounding block. The BLS estimate is the mean of the posterior distribution,  $p(\mathbf{x}|\mathbf{y})$ , which may be written in this case as:

$$\hat{\mathbf{x}}_{BLS}(\mathbf{y}) = \sum_{k=1}^K P(k|\mathbf{y}) \int_z \mathbb{E}\{\mathbf{x}|\mathbf{y}, k, z\} p_k(z|\mathbf{y}) dz.$$

The value of the integral over  $z$  is the BLS estimate for a single GSM model (as found in [7]), and the full estimate is a weighted sum of these over the  $K$  GSM mixture components. The integrals are relatively simple to compute since the embedded expectation is taken over a conditionally Gaussian density, and thus is just the Wiener solution,  $\mathbb{E}\{\mathbf{x}|\mathbf{y}, k, z\} = z\mathbf{C}_k(z\mathbf{C}_k + \mathbf{C}_w)^{-1}\mathbf{y}$ .

#### 3.2. Parameter estimation under additive noise

We adaptively estimate the parameters for each noisy subband, taking the set of all neighborhoods,  $\{\mathbf{y}_m; m = 1 \dots M\}$ , to represent a set of independent samples from the noise-corrupted MGSM density. We estimate the parameters that maximize the global log-likelihood expression of the observations:

$$L(\{\mathbf{y}_m\}) = \sum_{m=1}^M \log \left[ \sum_{i=1}^K P_{k_i} \int_z p(\mathbf{y}_m|k_i, z) p_{k_i}(z) dz \right]$$

In order to simplify the numerical optimization problem, we reduce the hidden multiplier densities  $p_k(z)$  to a finite number of discrete values. We maximize the likelihood using an iterative coordinate ascent method, optimizing each set of parameters in turn. The ascent steps are:

- *Discrete probabilities  $P_k$ :*

$$P_k^{(n+1)} = P_k^{(n)} \frac{1}{M} \sum_{m=1}^M \frac{p(\mathbf{y}_m|k)^{(n)}}{\sum_{j=1}^K p(\mathbf{y}_m|j)^{(n)} P_j^{(n)}}, \forall k \quad (2)$$

- *Mixing scale densities  $p_k(z)$ :*

$$p_k(z)^{(n+1)} = \frac{p_k(z)^{(n)}}{M P_k^{(n)}} \sum_{m=1}^M \frac{p(\mathbf{y}_m|k, z)^{(n)} P(k|\mathbf{y}_m)^{(n)}}{p(\mathbf{y}_m|k)^{(n)}}, \forall k \quad (3)$$

Note that, in contrast to original GSM model [7] where Jeffrey's noninformative prior was used for the mixing density, this model estimates the scale density directly from the data, as in [13].

- *Covariance matrices  $\mathbf{C}_k$ :* For the noisy case we have not found an efficient ML expression for  $\mathbf{C}_k$ . Instead we have used a consistent estimator which gives rise to the following updating rule:

$$\mathbf{C}_k^{(n)} = \left[ \left( \int_z \mathbf{C}_{\mathbf{y}}(k, z)^{(n)} p(z)^{(n)} dz \right) - \mathbf{C}_w \right]_+, \quad (4)$$

where the operator  $[\cdot]_+$  enforces positive definiteness by setting to zero the negative eigenvalues, and

$$\begin{aligned} \mathbf{C}_y(k, z)^{(n)} &= \frac{\sum_{m=1}^M P(k|\mathbf{y}_m)^{(n)} p(z|\mathbf{y}_m, k)^{(n)} \mathbf{y}_m \mathbf{y}_m^T}{\sum_{m=1}^M P(k|\mathbf{y}_m)^{(n)} p(z|\mathbf{y}_m, k)^{(n)}} \\ &\simeq z\mathbf{C}_k + \mathbf{C}_w. \end{aligned}$$

Without loss of generality, we have assumed that  $\int_z z p_k(z) dz = 1$  for all  $k$ , which leads directly to Eq. (4).

The ascent algorithm requires an initial set of parameter values. For both  $P_k$  and  $p_k(z)$  we assume a uniform distribution. The latter density is discretized using the sampling described in [7].

The covariances,  $\mathbf{C}_k$ , are initialized from the image content as follows: The first one,  $\mathbf{C}_1$ , is set to the global sample covariance. The remaining initial guesses for  $\mathbf{C}_k$ ,  $k = 2..K$  are obtained from high energy areas in the subband, selected using a heuristic.

To update the parameters accordingly, we need to compute, among others, the expression  $p(z|\mathbf{y}_m, k)$ . Using Bayes formula:

$$p(z|\mathbf{y}_m, k) = \frac{p(\mathbf{y}_m|k, z)p_k(z)}{p(\mathbf{y}_m|k)}.$$

This requires us to compute  $p(\mathbf{y}_m|k, z)$ , which is a zero-mean Gaussian distribution with covariance  $z\mathbf{C}_k + \mathbf{C}_w$ , and  $p(\mathbf{y}_m|k)$ , which is obtained by numerically integrating  $\int_z p(\mathbf{y}_m|k, z)p_k(z)dz$ . Both of these are involved in the update expressions, as well as  $P(k|\mathbf{y}_m)$ , which is also computed directly using Bayes formula.

## 4. RESULTS AND DISCUSSION

### 4.1. Implementation details

We decomposed the images using the Translation Invariant Haar Pyramid [8] (TIHP), with 3 orientations and 4 or 5 scales for  $256 \times 256$  or  $512 \times 512$  images respectively. We used  $K = 10$  GSMs in the mixture and, in contrast to [7], we increased the GSM neighborhood size from  $3 \times 3$  to  $7 \times 7$ . On each iteration of the parameter estimation algorithm, we pruned the number of GSMs in the mixture by classifying the samples according to  $P(k|\mathbf{y}_m)$  and deleting those GSMs with few samples assigned (less than  $N$ ).

We implemented the algorithm in *Matlab*. The current computational cost is quite high: roughly one hour for a  $256 \times 256$  image and six hours for a  $512 \times 512$  image, on a computer with two dual-core 3Ghz Intel Xeon processors and 32 Gbytes of memory. We believe a significant improvement should be possible with a more careful implementation, or by learning the covariance "offline", from a set of training images.

### 4.2. Denoising results

We used four standard test images to evaluate our algorithm, known as *Barbara*, *Boat*, *House* and *Peppers*. In Table 1 we present the results of our proposed algorithm, in terms of peak signal-to-noise ratio (PSNR) over a wide range of noise levels. Figure 2 shows a visual comparison for *Barbara* and *House*, two images with different amounts and types of texture, using simulated white Gaussian noise of  $\sigma = 25$ . Results are shown for the BLS-GSM method on TIHP (instead of full steerable pyramid, as in [7]), the SVGSM method [8], and the MGSM. All results are cropped to  $128 \times 128$  to promote the visibility of the artifacts, which are generally informative as to the relative advantages of denoising methods. We can see

$\sigma_w$ / PSNR	<i>Barbara</i>	<i>Boat</i>	<i>House</i>	<i>Peppers</i>
5/ 34.15	38.02	37.30	39.30	37.75
10/ 28.13	34.45	33.78	36.06	34.30
25/ 20.17	29.80	29.62	32.12	29.74
50/ 14.15	26.19	26.55	28.89	26.37
100/ 8.13	22.79	23.91	25.53	22.99

**Table 1.** Denoising performance expressed as peak signal-to-noise ratio (PSNR) in dB. First column shows the noise standard deviation and the PSNR of the noisy image.

the progressive and significant improvement (removing isolated basis functions, more natural edge appearance, better texture recovery, etc.) when introducing the spatial adaptation, and even more when using the mixture of GSMs.

Finally in Fig. 3 we show a graphical comparison of PSNR improvements (computed from the averaged MSE results of the 4 test images) with respect to the BLS-GSM method [7] on TIHP. Included are results for SVGSM [8], four of the current best published methods (labelled 3DBM [12], FoGSM [14], K-SVD [15], and SA-DCT [11]), and the MGSM model introduced here. The MGSM is seen to provide substantial improvement over the GSM and SVGSM, for all the images and noise levels. The average improvement is roughly 0.70 dB with respect to GSM and 0.40 dB with respect to SVGSM. Relative to other methods our results are better than SA-DCT and K-SVD, but they are still worse than FoGSM and 3DBM.

## 5. CONCLUSIONS AND FUTURE WORK

We have presented a denoising algorithm based on a mixture of Gaussian scale mixtures (MGSM). The proposed model provides a flexible and conceptually clean means of capturing non-local redundancy within subbands, while retaining the statistical strength of GSM-based models. In contrast to the original GSM and SVGSM models, there is no need for priors on the hidden scaling variables, as these are systematically ML-estimated for each subband. The proposed model leads to a denoising algorithm with consistent and significant improvement over GSM and SVGSM, and is competitive with methods proposed in recent literature.

We are currently investigating a variety of potential improvements. The pruning of under-utilized mixture components should be refined. The number of mixture models,  $K$ , should be selected adaptively per band, depending on both the content and the number of neighborhoods. And the model should be extended to exploit shared features across subbands, either by enlarging the GSM neighborhoods to include other bands, or sharing mixing variables across bands. We are also investigating more general issues of adaptive image representation that are inspired by this model.

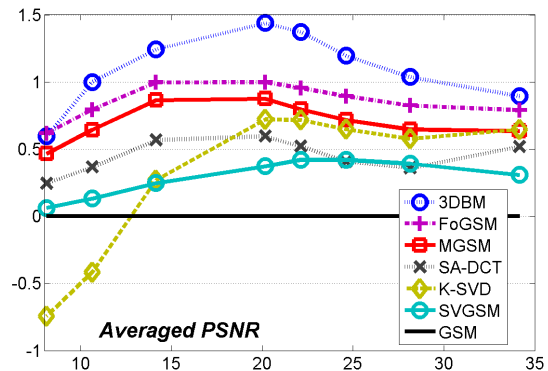
## 6. REFERENCES

- [1] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Comm.*, vol. COM-31, no. 4, pp. 532–540, April 1983.
- [2] 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.
- [3] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans Sig Proc*, vol. 41, no. 12, pp. 3445–3462, December 1993.



**Fig. 2.** Example denoised images. We compare three GSM-based methods, all implemented using TIHP, for *House* and *Barbara* images both with  $\sigma = 25$ . Images are cropped to  $128 \times 128$  pixels for visibility (starting in pixel [24,64] and [288,64] respectively [column, row]). For each  $2 \times 2$  panel, from left to right and top to bottom the PSNR values (*House / Barbara*) are: noisy observation (20.17/20.17); BLS-GSM method [7] (31.59/27.81); SVGSM [8] (31.71/28.81); our MGSM result (32.12/29.80).

- [4] 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.
- [5] D. Andrews and C. Mallows, "Scale mixtures of normal distributions," *J. Royal Stat. Soc.*, vol. 36, pp. 99–102, 1974.
- [6] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Adv. Neural*



**Fig. 3.** Comparison of denoising performance. Each curve shows PSNR improvement of one method, relative to GSM [7], as a function of input PSNR, averaged over 4 test images. Methods are: 3DBM [12]; FoGSM [14]; SA-DCT [11]; K-SVD [15]; GSM [7]; SVGSM [8] and our MGSM algorithm with  $K = 10$ . GSM, SVGSM, and MGSM are all implemented on a TIHP.

*Information Processing Systems (NIPS\*99)*, S. A. Solla, T. K. Leen, and K.-R. Müller, Eds., Cambridge, MA, May 2000, vol. 12, pp. 855–861, MIT Press.

- [7] 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.
- [8] J. A. Guerrero-Colón, L. Mancera, and J. Portilla, "Image restoration using space-variant Gaussian scale mixture in over-complete pyramids," *IEEE Trans. Image Proc.*, vol. 17, no. 1, pp. 27 – 41, January 2008.
- [9] 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.
- [10] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, July 2005.
- [11] 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.
- [12] 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.
- [13] J. Portilla, "Full blind denoising through noise covariance estimation using Gaussian scale mixtures in the wavelet domain," in *IEEE Int'l Conf. on Image Proc.*, IEEE, Ed. Singapore, October 2004, pp. 1217–1220.
- [14] S. Lyu and E. P. Simoncelli, "Statistical modeling of images with fields of Gaussian scale mixtures," in *Adv. Neural Information Processing Systems 19*, B Schölkopf, J Platt, and T Hofmann, Eds. May 2007, vol. 19, MIT Press.
- [15] 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.