A GEOMETRICAL A PRIORI FOR CAPTURING THE REGULARITY OF IMAGES

Yann Gousseau, and François Roueff

GET / Telecom Paris, CNRS LTCI
46 rue Barrault, 75634, Paris cedex, France
phone: + (33) 1 45 81 78 41, fax: + (33) 1 45 81 71 44, email: [gousseau,roueff]@tsi.enst.fr
web: www.tsi.enst.fr/~[gousseau,roueff]

ABSTRACT

We first briefly recall how to model occlusion and scaling in natural images through the use of a stochastic model, the scaling dead leaves model. Then we give a statistical estimator for its scaling parameters, which are related to the regularity of images. Last we show how this model can be used as an a priori for image denoising, in the framework of wavelet coefficients thresholding.

1. GENERAL CONTEXT

This communication is concerned with the regularity of natural images. It is well known that important features of natural images are scaling behaviour, strong discontinuities and relatively flat regions. These features are adequately modeled by a stochastic model involving the occlusion phenomena, the dead leaves model, when equipped with power law distributions ([13], [2], [8]). Roughly speaking, it may be seen as the result of the superimposing of objects with arbitrary scale. More recently, the existence of a limit at small scales for this model as well as its Besov regularity has been established ([7]). In this communication, we introduce an estimator for the scaling parameters of this model, therefore enabling to estimate its regularity. Then we investigate the use of such an estimation in the framework of denoising.

Two common approaches to wavelet based signal denoising are non parametric and Bayesian methods. The first ones usually rely on regularity assumptions and wavelet thresholding ([5]), whereas the second ones try to capture the specific shape of the distribution of wavelet coefficients for a given type of signal. Various type of priors have been considered for the marginals of wavelet coefficients. Being able to deal with the strongly non-Gaussian nature of these marginals, generalized Gaussian distributions have been widely used ([14]). Other priors include scale mixture of normal distributions ([12]) and Bessel K-forms ([6]). In Section 4, we investigate the use of the scaling dead leaves model to derive the specific shape of wavelet marginals to be used in a bayesian approach. Wavelet methods do not allow to take full advantage of the geometry of the model especially since we will only focus on marginals. Nevertheless, as shown in [1] (in the framework of non-parametric denoising), specifying an *a priori* on wavelet marginals can be related to regularity assumptions. Here, for similar reasons, the wavelet *a priori* will inherit the regularity of the model as established in [7].

2. A MODEL FOR REGULARITY AND ITS WAVELET COEFFICIENTS

In this section, we sketch the construction of a model for natural images, omitting technical details. The interested reader
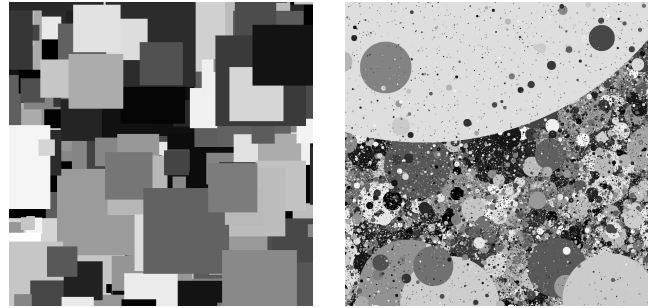

Figure 1: Left : a dead leaves model constructed from squares with a uniform distribution. Right : the limit at small scales of a dead leaves model with size distribution (1), $\alpha = 2.9$.

is referred to [7]. We first consider a family of independent and identically distributed (i.i.d.) random objects : compact subsets X_i of $\mathbb{R}^2$, equipped with random variables U_i, their gray levels. For the sake of simplicity, we will assume that the U_i are uniformly distributed. Then, we consider the random image that is obtained when these objects are put at random positions in the plane, one by one, and such that any object may hide previously laid ones if it happens to be on the top of it. To be more precise, we consider the limit of the Markov chain obtained this way. Figure 1, left, illustrate such a construction when the objects are all squares with a uniformly distributed radius. This simple model, first introduced by G. Matheron ([11]), has been shown to be adequate for image modeling when objects have a power law distribution of their size ([13], [2], [8]), that is when $X_i = R.Y$, R being a random variable with density

$$f(r) \quad r^{-\alpha}. \tag{1}$$

The main interest of this model is its ability to handle many statistics of natural images with few parameters (see the aforementioned works). However, Equation (1) imposes small and large sizes' cut-offs, respectively r_0 and r_1, for the model to be well defined. In particular, the minimum objects' size, parameter r_0, influences the small scale regularity of the model and would obviously need to be taken into account within a denoising procedure. However, when $\alpha < 3$ (a reasonable range to model natural images, as will be seen later) it may be shown (see [7]) that there exists a limit model I_0 when $r_0 \to 0$. In what follows, we will refer to I_0 as the Scaling Dead Leaves model (SDL). This model exhibits interesting geometrical structures, as may be seen from the two dimensional distribution of I_0. Indeed, the model is stationary and, for any $x \in \mathbb{R}^2$, $(I_0(0), I(x))$ is distributed as a mix-

ture of $(\mathscr{U},\mathscr{U})$ and $(\mathscr{U},\mathscr{U}')$, where $\mathscr{U}$ and $\mathscr{U}'$ are independent uniform variables, with respective weights $p(x)$ and $1-p(x)$. The interpretation of this mixture is that two points at distance x will be in the same objects with probability $p(x)$ and in different independent objects otherwise. Even though the limit model is not piecewise constant (in particular there are small objects everywhere), this distribution reflects the presence of weakly uniform regions, as may be seen on Figure 1, right. Moreover, it may be shown that

$$p(x) \sim 1 - Cx^{3-\alpha} \text{ as } x \to 0, \tag{2}$$

where C is a constant depending on the geometry of objects X_i. From this mixture, it is easily seen that the distribution of "jumps" in the image, $I_0(x) - I_0(0)$, consists of the mixture of a point mass at 0 and a symmetrical distribution corresponding to the difference between to uniformly and independently distributed uniform variables. This distribution is in agreement with the general shape observed for the distribution of such jumps in an image, highly peaked at 0 and slowly decreasing away from zero towards the extreme values. This behaviour together with the power law in (2) inherited from the distribution of the size of objects (1) allows to show that the Besov regularity of I_0 is mainly driven by the parameter α, see [7].

In Section 4, in the framework of wavelet-based image denoising, we approximate the distribution of coefficients with this same mixture, therefore neglecting an averaging phenomenon for which it is difficult to derive analytic formulas. We will see in the experiments a drawback of this approximation. Let $w_{k,j}^{\eta}$ be the wavelet coefficients of the image (in L^2 normalization), η being the orientation, k the location and j the octave. Let us specify that in our notation, $j = 0$ corresponds to the entire image and that as j increases, one goes toward the pixel size. We assume that

$$w_{k,j}^{\eta} \sim g_j(w)(\delta_0 + \lambda)(dw), \tag{3}$$

where δ_0 is the Dirac mass at 0, λ the Lebesgue measure, and $g_j(w)$ denotes the *a priori* density of $w_{k,j}^{\eta}$ with respect to the measure $\delta_0 + \lambda$, and is defined for all w as

$$p_j \mathbf{1}(w = 0) + (1 - p_j)M_j^{-1}(1 - |w|/M_j)_+ \mathbf{1}(w \neq 0), \tag{4}$$

where $y_+ := y \vee 0$, $p_j = p(2^{-j})$ and $M_j = M2^{j-1}$, where M is the maximum of gray level of the model I_0 (so that M_j is the theoretical maximum of wavelet coefficients at octave j). Although the above mixture, (3), is an idealization, it is easily seen that it yields a correct formula for the variance of the wavelet coefficients of the model up to a multiplicative constant independent of the octave j. This last point is to be used in Section 3.

3. ESTIMATION OF THE SCALING PARAMETERS

In this section, we define an estimator for the parameters α and C. The first interest of such an estimator is to provide a parametric estimation of Besov regularity. This is an alternative to the method presented in [4], based on compression performances. This estimator will be validated on simulations of the SDL model. Our second interest is related to Bayesian image denoising. In order to compute the MAP estimator $\hat{w}_{k,j}^{\eta}$, one needs to estimate α and C of the *a priori*

model, the so-called *hyper-parameters*. On most natural images we have observed that α ranges between 2 and 3, values which are in agreement with the theoretical model (see [7]). Note that $\alpha \sim 3$ corresponds to highly textured images while $\alpha \sim 2$ corresponds to more flat images. Here we propose an estimator based on a contrast computed from the empirical variance of the wavelet coefficients. Indeed, from (3) and (6), we compute (recall that σ^2 is assumed to be known)

$$\text{Var}(x_{k,j}^{\eta}) = C2^{j(\alpha-3)}M_j^2/6 + \sigma^2 =: v_j(\alpha,C).$$

Observe also that the coefficients have zero mean. Define

$$\mathscr{C}(\alpha,C) = \sum_{j,k,\eta} \frac{x_{k,j}^{\eta\,2}}{v_j(\alpha,C)} + \sum_{j,k,\eta} \log v_j(\alpha,C),$$

where the two sums are computed for all available η, j, k up to some cut-off $j \leq J$. We then define the contrast estimator

$$(\hat{\alpha},\hat{C}) := \underset{\alpha,C}{\arg\min} \, \mathscr{C}(\alpha,C). \tag{5}$$

Let us briefly justify why $\mathscr{C}$ can be interpreted as a contrast, yielding a reasonable estimator $(\hat{\alpha},\hat{C})$ (although we do not provide any proof). Let α^* and C^* denote the true parameters. As the number of wavelet coefficients increases, up to some renormalizing factor, $\mathscr{C}(\alpha,C)$ behaves as its expectation

$$\mathbb{E}\mathscr{C}(\alpha,C) = \sum_{j,k,\eta} \frac{v_j(\alpha^*,C^*)}{v_j(\alpha,C)} + \sum_{j,k,\eta} \log v_j(\alpha,C).$$

Now observe that $\mathbb{E}\mathscr{C}(\alpha,C) - \mathbb{E}\mathscr{C}(\alpha^*,C^*)$ equals

$$\sum_{j,k,\eta} \left[\left(\frac{v_j(\alpha^*,C^*)}{v_j(\alpha,C)} - 1 \right) - \log \frac{v_j(\alpha^*,C^*)}{v_j(\alpha,C)} \right]$$

which is minimized at $\alpha = \alpha^*$ and $C = C^*$ since $x - 1 - \log(x)$ is minimized at $x = 1$ over $x > 0$. For computing (5), we used a standard gradient minimization approach within a bounded domain for (α,C). Experiments have shown quite robust estimates even in the presence of noise. The function $\mathscr{C}(\alpha,C)$ is represented in Figure 2, left, for the Lena image with additive Gaussian noise ($\sigma = 10$).

4. BAYESIAN WAVELETS THRESHOLDING

Suppose that we observe an image polluted by an independent Gaussian additive white noise with known variance σ^2. Let us denote by $w_{k,j}^{\eta}$ its true wavelet coefficients and by $n_{k,j}^{\eta}$ the wavelet coefficients of the additive noise. We let

$$x_{k,j}^{\eta} := w_{k,j}^{\eta} + n_{k,j}^{\eta} \tag{6}$$

denote the observed wavelet coefficients. We adopt a Bayesian approach by defining $\hat{w}_{k,j}^{\eta}$ as the MAP (maximum *a posteriori*) estimator corresponding to the *a priori* defined as follows; $\{w_{k,j}^{\eta}, \eta, j, k\}$ are independent and their marginal distributions are given by (3). Observe that the independence is not a property of the model introduced in Section 2 but a technical simplification for deriving a simple denoising procedure. In particular, using this independence assumption,
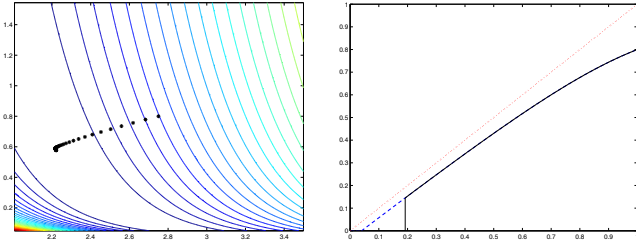
Figure 2: Left: $\mathscr{C}(\alpha,C)$ (level curves) on a grid $\alpha \in (2,3.5)$ (horizontal) and $C \in (0.05,1.55)$ (vertical). The $*$'s represent the trajectory of a gradient descent (on the well-known Lena image). Right: The Bayesian threshold function T_j with $\alpha = 2.8$, $C = 1.0$, $\sigma/M = 0.2$ and $j = 8$; horizontal axis: x/M, vertical axis: $T_j(x)/M$ (plain), $T_j^{(0)}(x)/M$ (dashes) and x/M (dots).

the resulting MAP estimator consists in thresholding the coefficients $x_{k,j}^\eta$, namely

$$\hat{w}_{k,j}^\eta = \arg\max_w f_j(w|x_{k,j}^\eta) =: T_j(x_{k,j}^\eta) , \qquad (7)$$

where $f_j(\cdot|x)$ denotes the conditional density of $w_{k,j}^\eta$ given $x_{k,j}^\eta = x$ with respect to the measure $\delta_0 + \lambda$. Let us now compute the resulting thresholding function T_j. Using usual conditional density formula, since that given $w_{k,j}^\eta = w$, $x_{k,j}^\eta$ is a $\mathscr{N}(w,\sigma^2)$ r.v., we have that

$$f_j(w|x) \quad \exp\left(-\frac{(x-w)^2}{2\sigma^2}\right) g_j(w) . \qquad (8)$$

Define $T_j^{(0)}(x) := \arg\max_{w\neq 0} f_j(w|x)$. Using (4), straightforward computations then yield

$$T_j^{(0)}(x) = \frac{\text{sign}(x)}{2}\left[M_j + |x| - \sqrt{(M_j - |x|)^2 + 4\sigma^2}\right]_+ \quad (9)$$

$$T_j(x) = T_j^{(0)}(x)\,\mathbf{1}\left\{f_j(T_j^{(0)}(x)|x) > f_j(0|x)\right\} . \qquad (10)$$

As may be seen in Figure 2, right, T_j is similar to a hard threshold function. This follows from the point mass component of the *a priori* distribution. Therefore it will suffer from the same drawbacks as hard thresholding denoising (see [10]). It is also interesting to note that the true wavelet coefficients distribution of the SDL model, whose precise computation is an open question, would not have such a point mass component, and would lead to a soft-like thresholding.

Collecting (4), (7), (8), (9) and (10) allows to compute $\hat{w}_{k,j}^\eta$ from $x_{k,j}^\eta$ in a few simple computations.

5. EXPERIMENTAL RESULTS

As a first experiment, we validate the estimator of Section 3 on simulations of the SDL model for two different values of α (2.3 and 2.8) and various noise levels, see Figure 3. In these simulations, objects are disks with distribution given by (1), $r_0 = 1$, $r_1 = 2048$ and images are of size 1024×1024. From this experiment, we see that the procedure is quite robust to noise, and that α is correctly estimated with a bias of about $+0.1$.
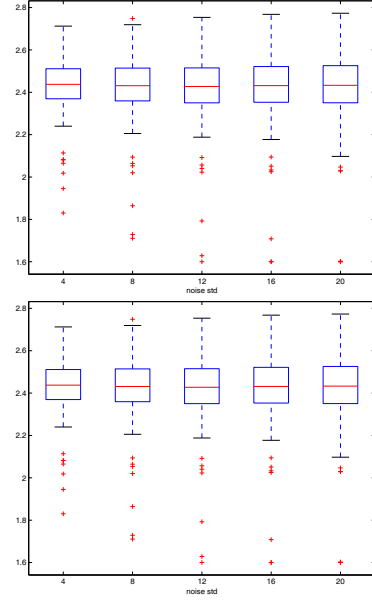

Figure 3: Boxplots for the estimation of α on synthetic images. Top : $\alpha = 2.3$, bottom : $\alpha = 2.8$.
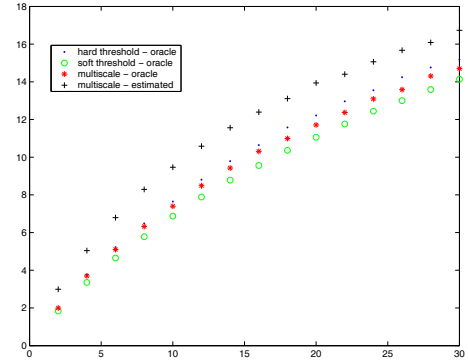

Figure 4: Comparison of the averaged $L2$ errors of the oracle methods for hard, soft and multi-scale, and of the multi-scale method with estimated coefficients (Lena image).

Next we compare the denoising with a SDL *a priori* with two other usual wavelet thresholdings, namely the so-called hard and soft thresholdings (see [10]), using Daubechies wavelets of order 3. We first compare oracle performances. This means that we compare the lowest errors obtained over a large sample of hyper-parameters (α, C) or thresholds t. We obtain slightly lower errors for our method than with hard thresholding (see Figure 4, obtained for the Lena image) at large noise variances. This is easily explained by the fact that we optimize the error over a two-dimensional parameter in our method and a one-dimensional one for the usual thresholdings and this is more effective for strong noise as more scales are involved in denoising. However, this is no longer true for the comparison with the soft thresholding oracle, see also the discussion at the end of Section 4.

Oracle estimations are not realistic in practice as hyperparameters are computed by using the original image which is not at hand. On the same figure, 4, we also display

Figure 5: Top left : original image; top right : SDL denoising (translation invariant) after adding Gaussian noise ($\sigma = 10$); bottom right : detail of SDL denoising, bottom right : detail of oracle hard threshold denoising (translation invariant).

the performance of the proposed method when the hyper-parameters are estimated as in Section 3. Of course results are slightly degraded, but stay reasonably close. Finally we provide a visual comparison of multi-scale (with estimated hyper-parameters) and oracle hard thresholding on a high resolution image in Figure 5 (note that images may be extracted from the pdf file of this paper, e.g. by using the "pdfimages" command on Linux systems). We used translation invariant wavelet transform for a better visual result, a usual additional step of wavelet denoising methods (see [10]). Parameters are estimated as explained in Section 3. On this figure, we can see that the SDL *a priori* allows to better preserve the presence of flat zones than the hard threshold, which result in less artefacts. This will also be the case with a multi-sure soft thresholding method.

6. CONCLUSION

In this communication, we introduced an estimator for scaling parameters of a geometrical model for natural images. Then we investigated the use of such a geometrical *a priori* in the context of denoising. This approach gives raise to a Bayesian thresholding method along the same lines as recent ones based on empirical studies of wavelet coefficient statistics. However, in the latter approaches, parameters of the Bayesian *a priori* have to be estimated at each scale whereas we only need to specify two parameters for all scales. From a practical point of view, this allow more robust estimations of hyper-parameters, in particular at scales (the more important ones for denoising) where the signal/noise ratio is low.

We do not claim to achieve results better to state of the art methods, and one of the aims of this paper is to provide a sanity check for the model presented in [7] in the framework of denoising. In particular, we do not take into account dependences produced by discontinuities or textures, better

dealt with using methods such as [3] or [9]. Possible extensions of our approach could be to use over-complete bases to better preserve textures.

From a theoretical point of view, the studied estimation problem is important since, as in [1], the hyper-parameters α and C can be related to the regularity of the model (established in [7]). In other words, although the statistical *a priori* is limited to wavelet marginals, the regularity properties are inherited from the geometrical model and therefore bears some physical justifications. An interesting perspective could be to compare the regularity estimated through compression performances, as in [4], with the estimation procedure presented in this paper.

REFERENCES

[1] F. Abramovich, T. Sapatinas, and B. W. Silverman. Wavelet thresholding via a Bayesian approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(4):725–749, 1998.

[2] L. Alvarez, Y. Gousseau, and J.-M. Morel. The size of objects in natural and artificial images. *Advances in Imaging and Electron Physics, Academic Press*, 111:167–242, 1999.

[3] T. Buades, T. Coll, and J.-M. Morel. On image denoising methods. Technical report, CMLA, 2004.

[4] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Transactions on information theory*, 38(2), 1992.

[5] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. *Ann. Statist.*, 26(3):879–921, 1998.

[6] J. M. Fadili and L. Boubchir. Analytical form for a bayesian wavelet estimator of images using the bessel K form densities. *IEEE Trans. Image Processing*, 14(2):231, 2005.

[7] Y. Gousseau and F. Roueff. Modeling occlusion and scaling in natural images. Submitted, 2004.

[8] A. Lee, D. Mumford, and J. Huang. Occlusion models for natural images: A statistical study of a scale invariant dead leaves model. *Int'l J. of Computer Vision*, 41:35–59, 2001.

[9] F. Malgouyres. A noise selection approach of image restoration. In *SPIE, Int. Conf. on Wavelets IX*, volume 4478, pages 34–41, San Diego, 2001.

[10] S Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 2nd edition, 1999.

[11] G. Matheron. Modèle séquentiel de partition aléatoire. Technical report, CMM, 1968.

[12] J Portilla, V Strela, M Wainwright, and E P Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Image Processing*, 12(11):1338–1351, 2003.

[13] D. L. Ruderman. Origins of scaling in natural images. *Vision Research*, 37(23):3385–3398, 1997.

[14] E. P. Simoncelli. Statistical model for images: Compression, restoration and synthesis. In *31st Asilomar Conference on Signal, Systems and Computers*, 1997.