14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

A PERCEPTUAL BAYESIAN ESTIMATION FRAMEWORK AND ITS APPLICATION TO IMAGE DENOISING

Javier Portilla

Visual Information Processing Group, DECSAI
Universidad de Granada, Spain
javier@decsai.ugr.es, http://decsai.ugr.es/∼javier

ABSTRACT

We present a generic Bayesian framework for signal estimation that incorporates into the cost function a perceptual metric. We apply this framework to image denoising, considering additive noise of known density. Under certain assumptions on the way local differences in visual responses add up into a global perceptual distance, we obtain analytical solutions that exhibit interesting theoretical properties. We demonstrate through simulations, using an *infomax* nonlinear perceptual mapping of the input and a local Gaussian model, that in the absence of a prior the new solutions provide a significant improvement on the visual quality of the estimation. Furthermore, they also improve in Mean Square Error terms w.r.t. their non-perceptual counterparts.

1. INTRODUCTION

The Bayesian approach provides a powerful framework for signal estimation. Besides the primary source of information (i.e., the observation itself), it considers three more sources: (1) the degradation model; (2) the prior model; and (3) the error-cost model. When looking at the modern literature on non-blind image denoising, we see that most often the degradation model is assumed known, and, in fact, performance tests are typically made through simulating the degradation of a set of originals. As such, these experiments do not measure the realism or accuracy of the degradation model. The room for the estimation improvement, instead, is mainly left to the second additional source of information: the prior on the uncorrupted signal. We have observed during the last two decades that most significant advances in image denoising have been achieved by using more powerful prior statistical models. And not as much by increasing the prior sophistication, as by using new image representations, such as wavelets or overcomplete pyramids, that allow for a simpler expression of key statistical features of the images [1].

Then, what has been the role of the third additional source of information for image estimation, i.e., the error-cost function? In the important case when the final destination of the estimation is the human visual system (HVS), then we clearly have to consider the perceptual impact of the estimation error in the cost function design [1]. Unlike in other fields, such as image and video compression, where some simplified perceptual models are being used for the last 10 years or more (e.g. [2]), in image estimation little attention has been paid to perceptual issues. A significant fact is that performance is still measured, in the overwhelming majority of cases, by the mean square error (MSE), when there is general agreement about the inadequacy of this index as an image fidelity measurement. Some more powerful alternative measurements have been proposed (e.g. [3, 4]). However, till this date and to the best of our knowledge, no attempt has been made to include any perceptual measurement as an error-cost function to be minimized within a Bayesian estimation frame.

In this paper we propose a way to integrate the available perceptual and statistical information into a single Bayesian estimation frame. Although we will make some assumptions that are based on mathematical convenience, the resulting frame is general enough (1) to deal with an interesting variety of situations; (2) to give some insight about the problem; and (3) to produce some encouraging preliminary results. Being this our first attempt to add perceptual and statistical information into a single estimation, we have not tried to use here a sophisticated or realistic perceptual model. Instead, we have used the *info-max* concept [5] to derive simplified models able to reproduce visual phenomena like saturation and masking. Note that by using a statistical criterion to derive a perceptual model we are implicitly following the efficient encoding conjecture [6], which consists of interpreting the low-level internal visual representation as a maximum-entropy mapping of the visual stimuli, i.e., a non-linear multidimensional transform designed to maximize, in statistical terms, the information capacity of the HVS (and, therefore, to minimize the mutual information among the responses). Although the referred conjecture is by no means accepted as an ultimate truth (see, e.g. [7, 8]), it has consistently demonstrated over the last years that it is a powerful approach for deriving basic properties, both linear and non-linear, of the early vision (see, e.g. [11, 7, 9]).

2. A UNIFIED FRAMEWORK FOR PERCEPTUAL-STATISTICAL ESTIMATION

Consider an original vector $\mathbf{x}$ subject to a degradation described by $p(\mathbf{y}|\mathbf{x})$, where $\mathbf{y}$ is the observation, and assume a prior statistical knowledge on the original given by $p(\mathbf{x})$. Assume also that we have a perceptual metric, i.e., a function $d_p : R^N \times R^N \longrightarrow R^*$ that gives a non-negative measurement of how perceptually distant are two vectors. We define our Perceptual Bayes Least Squares (PBLS) estimator as:

$$\hat{\mathbf{x}}_{PBLS} = \arg\min_{\xi} \mathbb{E}\{d_p^2(\mathbf{x},\xi)|\mathbf{y}\} = \arg\min_{\xi} \int_{\mathbf{x}} p(\mathbf{x}|\mathbf{y})\; d_p^2(\mathbf{x},\xi) d\mathbf{x}. \quad (1)$$

This work is supported by the Ministerio de Educacion y Ciencia (MEC, Spain) through grant TIC2003-01504 and the ”Ramon y Cajal Program”.

[1]The results presented here are also applicable to audio signals.

This estimation provides a *perceptual representative* of the posterior pdf, for a given observation $\mathbf{y}$. However, in practical cases we do not completely know d_p. Instead, we may have a model of how our perceptual system responds to the stimuli. Ideally, such a model would provide a transformation of the original stimulus to a certain *perceptual domain* in which equally distant vectors (according to the Euclidean metric) would correspond to perceptually equidistant pairs of stimuli (e.g., [10]). This transformation is a (non-linear) mapping $\mathbf{f} : R^N \longrightarrow \mathscr{D}_\mathbf{f} \subset R^M$, with $M \geq N$, accomplishing

$$d_p(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\|,$$

and Eq. 1 can be rewritten as:

$$\hat{\mathbf{x}}_{PBLS} = \arg\min_{\xi} \int_{\mathbf{x}} p(\mathbf{x}|\mathbf{y})\ \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\xi)\|^2 d\mathbf{x}. \quad (2)$$

We also impose that if $\mathbf{f}(\mathbf{x}_1) = \mathbf{f}(\mathbf{x}_2)$ then $\mathbf{x}_1 = \mathbf{x}_2$, i.e., $\mathbf{f}$ is invertible. Expressing $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\xi)\|^2$ as $\|\mathbf{f}(\mathbf{x})\|^2 + \|\mathbf{f}(\xi)\|^2 - 2\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\xi)$, and calling C the integral of Eq. 2, we obtain $C = \mathbb{E}\{\|\mathbf{f}(\mathbf{x})\|^2|\mathbf{y}\} + \|\mathbf{f}(\xi)\|^2 - 2\mathbf{f}(\xi) \cdot \mathbb{E}\{\mathbf{f}(\mathbf{x})|\mathbf{y}\}$. We differentiate C w.r.t. each component of ξ, resulting:

$$\frac{\partial C}{\partial \xi_i} = 2\frac{\partial \mathbf{f}(\xi)}{\partial \xi_i} \cdot (\mathbf{f}(\xi) - \mathbb{E}\{\mathbf{f}(\mathbf{x})|\mathbf{y}\}) = 0. \quad (3)$$

When $M = N$ the previous equation can be expressed in terms of the Jacobian of $\mathbf{f}$, $\mathbf{J}_\mathbf{f}(\mathbf{x})$, as

$$\mathbf{J}_\mathbf{f}(\mathbf{x})\,(\mathbf{f}(\xi) - \mathbb{E}\{\mathbf{f}(\mathbf{x})|\mathbf{y}\}) = \mathbf{0}.$$

Because $\mathbf{f}$ is invertible, $\det(\mathbf{J}_\mathbf{f}(\mathbf{x})) \neq 0$, and the solution is

$$\hat{\mathbf{x}}_{PBLS} = \mathbf{f}^{-1}(\mathbb{E}\{\mathbf{f}(\mathbf{x})|y\}). \quad (4)$$

When $M > N$ this latter solution would also be admissible if $\mathbb{E}\{\mathbf{f}(\mathbf{x})|y\} \in \mathscr{D}_\mathbf{f}$. We could even drop the requirement of $\mathbf{f}$ being invertible, whenever, besides the latter condition, there was a criterion for choosing among multiple solutions of Eq. 3 (e.g., minimum L2 norm). Note that when $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ $\hat{\mathbf{x}}_{PBLS}$ becomes the Minimum Mean Square Error estimator.

3. ESTIMATION UNDER ADDITIVE INDEPENDENT NOISE AND NO PRIOR

Consider now that:

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

where $\mathbf{w}$ represents noise, whose pdf, p_w is known. Let's assume also that there is no prior knowledge on $\mathbf{x}$. Then $p(\mathbf{x}|\mathbf{y}) = p_w(\mathbf{y} - \mathbf{x})$. Consider now a single component of $\mathbf{f}(\mathbf{x})$, $f_j(\mathbf{x})$. According to Eq. 4 we can write:

$$\begin{aligned} f_j(\hat{\mathbf{x}}_{PBLS}) &= \int_{R^N} p_w(\mathbf{y} - \mathbf{x}) f_j(\mathbf{x})\, d\mathbf{x} \\ &= (f_j * p_w)(\mathbf{y}), \end{aligned} \quad (5)$$

where symbol "*" represents convolution. Then, through a slight abuse of notation, we can gather all vector components back into $\mathbf{f}(\hat{\mathbf{x}}_{PBLS}) = (\mathbf{f} * p_w)(\mathbf{y})$, and inverting:

$$\hat{\mathbf{x}}_{PBLS} = \mathbf{f}^{-1}\left[(\mathbf{f} * p_w)(\mathbf{y})\right]. \quad (6)$$

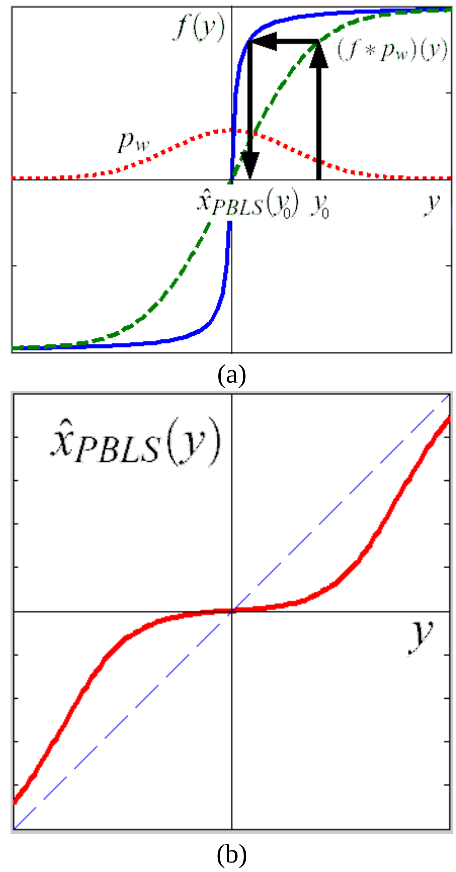

Figure 1: (a) Graphical explanation of Eq. 6. Dot line is the noise pdf p_w, solid line is $f(y)$ and dash line is $(f * p_w)(y)$. (b) The resulting shrinkage function $\hat{x}_{PBLS}(y)$, for the function and noise pdf of panel (a).

This result is very interesting: the estimation is the input that would give raise to a local average (through convolution with the noise density) of the perceptual responses around the observation. As shown in Fig. 1, under typical noise pdf's and perceptual functions (symmetric compressive nonlinearities), the effect is to shrink the observations with small amplitude, very much like for classical Bayesian coring (e.g., [12]).

3.1 Infomax mapping

From here on we consider that the non-linear mapping aims to achieve the maximum entropy representation of the uncorrupted signal, subject to a given output interval. This is related, on the one hand, with Barlow's efficient encoding conjecture [6], because maximizing the joint entropy subject to some representation constraints implies minimizing the mutual information among neural responses. On the other hand, it is related to the "information maximization" principle proposed in Ref. [5], which aims to separate the independent components of the input. Although proposed in a different context and for a different (though related) goal, we will also use here the term *infomax*, but now to refer to a choice of the perceptual mapping that aims to maximize the entropy of the internal representation of the original (uncorrupted) signals.

We consider first only marginal statistics. According to the *infomax* criterion, and choosing $(0,1)$ for the output interval [2], our non-linear mapping $f_x(x)$ is the cumulative density function of the original:

$$f_x(x) = \int_{-\infty}^{x} p_x(x')dx'. \quad (7)$$

This function forces $f_x(x)$ to be uniformly distributed in $(0,1)$ when $x \sim p_x$. Note also that, as long as $p_x(x)$ decreases in modulus as $|x|$ increases (something that typically happens after having band-pass linear filtered the data), $f_x(x)$ reduces its sensitivity (its derivative) as we increase the magnitude of the input, until reaching a saturation level. This is a typical behavior in biological perception.

As no prior is considered for the Bayes formula [3], we have $p(x|y) = p_w(y-x)$. For this case, and according to Eq. 5, $f_x(\hat{x}_{PBLS})$ can be computed by:

$$\mathbb{E}\{f_x(x)|y\} = (f * p_w)(y).$$

By defining the step function $u(x)$ as 0 for $x<0$ and 1 for $x \geq 0$, we can write $f_x(x)$ (Eq. 7) as $u(x) * p_x(x)$. Then,

$$\begin{aligned}\mathbb{E}\{f_x(x)|y\} &= (u * p_x * p_w)(y) \\ &= (u * p_y)(y) \\ &= \int_{-\infty}^{y} p_y(y')dy',\end{aligned}$$

where we have applied that, because of being x and w independent, the marginal pdf of the observation is $p_y = p_x * p_w$. Here the r.h.s. is the cumulative density function of the observation y, f_y. Therefore, $\mathbb{E}\{f_x(x)|y\} \sim u(0,1)$ and, as a consequence,

$$\hat{x}_{PBLS}(y) = f_x^{-1}(f_y(y)) \sim p_x.$$

This estimator is a non-linear operation that transform point-wise the observation to impose the original (uncorrupted) marginal statistics. This is a remarkable property, because usually the estimators have quite a different statistical distribution than the uncorrupted signal.

It is interesting to note that, if we were considering a prior also in the posterior calculation (which should increase the estimation quality), by following this *infomax* adaptation paradigm the statistics of the original signal would be considered *twice* in the estimation: once in computing the posterior, and once again through the non-linear response function.

4. A LOCAL INFOMAX WIENER ESTIMATE IN AN OVERCOMPLETE ORIENTED PYRAMID

4.1 Infomax Wiener estimate

Let's assume now that both our signal and noise are modelled marginally as zero-mean Gaussian of known variance (σ_x^2 and σ_w^2, respectively). As before, we consider our non-linear response adapted to represent with maximum entropy the uncorrupted signal. We have used the definition $erf(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} exp(-0.5(x')^2)dx'$, and we use now $f(x) = erf(x/\sigma_p)$ for the *perceptual* mapping. Although we would choose $\sigma_p = \sigma_x$ for an uniform response of the uncorrupted signal (*infomax* solution), we intentionally leave σ_p as a free parameter, so we can study how its choice affects to the final solution.

According to Eq. 4 and to our choice for $f(x)$ we can write:

$$f(\hat{x}_{PBLS}) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} \frac{\exp\left(-\frac{(x')^2}{2\sigma_p^2}\right)}{\sqrt{2\pi\sigma_p^2}}dx' \frac{\exp\left(-\frac{(x-\hat{x}_W(y))^2}{2\hat{\sigma}_W^2}\right)}{\sqrt{2\pi\hat{\sigma}_W^2}}dx, \quad (8)$$

where the inner integral is $f(x)$ and the factor outside the inner integral is the Gaussian posterior given by the Wiener solution, with mean and variance given by:

$$\begin{aligned}\hat{x}_W(y) &= \frac{\sigma_x^2}{\sigma_x^2+\sigma_w^2}y \\ \hat{\sigma}_W^2 &= \frac{\sigma_w^2\sigma_x^2}{\sigma_x^2+\sigma_w^2}.\end{aligned}$$

First we note that Eq. 8 can be written as a convolution of $f(x)$ with a Gaussian kernel. As before, we also note that $f(x)$ itself can be also written as the convolution of another Gaussian kernel with the integration kernel $u(x)$. Therefore, we can write $f(\hat{x}_{PBLS})$ as an integral of the convolution of two Gaussian functions, or, in other words, as the integral of a Gaussian whose variance is the sum of the two convolved Gaussian functions:

$$\begin{aligned}f(\hat{x}_{PBLS}) &= (f * g_W)(\hat{x}_W(y)) \\ &= ((u * g_p) * g_W)(\hat{x}_W(y)) \\ &= (u * (g_p * g_W))(\hat{x}_W(y)) \\ &= (u * g_s)(\hat{x}_W(y)),\end{aligned}$$

where g_p, g_W and g_s are zero-mean Gaussian pdf's with variances σ_p^2, $\hat{\sigma}_W^2$, and $\sigma_s^2 = \sigma_p^2 + \hat{\sigma}_W^2$, respectively. Applying the definition of $f(x) = erf(x/\sigma_p)$ and the previous result we can re-express the previous equation as:

$$\int_{-\infty}^{\hat{x}_{PBLS}} \frac{\exp\left(-\frac{(x')^2}{2\sigma_p^2}\right)}{\sqrt{2\pi\sigma_p^2}}dx' = \int_{-\infty}^{\hat{x}_W} \frac{\exp\left(-\frac{x^2}{2\sigma_s^2}\right)}{\sqrt{2\pi\sigma_s^2}}dx.$$

As both integrals (left and right side) are identical except for a scale factor and the upper integration limit, it must be the case that $x' = (\sigma_p/\sigma_s)x$, which implies in turn that

$$\hat{x}_{PBLS} = (\sigma_p/\sigma_s)\hat{x}_W(y).$$

Substituting the values of σ_s, $\hat{x}_W(y)$ and $\hat{\sigma}_W$, and operating, we finally obtain:

$$\hat{x}_{PBLS} = \frac{\sigma_x^2}{\sqrt{(\sigma_x^2+\sigma_w^2)(\sigma_x^2+\sigma_w^2(1+\sigma_x^2/\sigma_p^2))}}y. \quad (9)$$

In particular, for the *infomax* case we have $\sigma_p = \sigma_x$ and $\hat{x}_{PBLS} = \frac{\sigma_x^2}{\sqrt{(\sigma_x^2+\sigma_w^2)(\sigma_x^2+2\sigma_w^2)}}y$. This new estimator *shrinks* more the signal than the classical Wiener, but is still linear.

[2] We will use here the interval (0,1) for simplicity, but a more convenient mapping keeping the 0 to 0 correspondence (e.g. to the interval (-1,1)), could be used as well. Both mappings provide exactly the same estimation, because Eq. 4 is not affected by any affine transform of $\mathbf{f}(\mathbf{x})$ (only its non-linear component has an effect on the solution).

[3] Using the prior information for adapting the non-linearity to the data, but not for the posterior computation may seem unusual, but it is as legitimate as using it for the posterior, but not for the non-linear adaptation, as usually done.

4.2 Local adaptation

The solution described in previous subsection has, by itself, a limited estimation power. First, it only deals with marginal statistics. Recent successful results in denoising come from using a neighborhood around each image coefficient for the estimation of the central or reference coefficient (e.g. [13, 14, 15]. Secondly, Gaussian models, either joint or marginal, are clearly insufficient for image modelling when applied to the image globally (e.g. [1]). However, high-performance denoising can be obtained by considering that the uncorrupted signal (e.g., in an overcomplete pyramid representation) is *locally* Gaussian. An effective procedure, then, is to estimate the variance of the local Gaussian in a neighborhood, and then apply the Wiener solution locally to the central coefficient, subtracting the variance of the noise to the local sample variance [16, 17, 18]. Underneath is the local image modelling in the transformed domain as a Gaussian Scale Mixture, i.e., as the product of a known zero-mean Gaussian vector times a hidden positive scalar [19, 20, 15]. Following the *infomax* criterion, we will apply an adaptive non-linear function to maps every variable Gaussian density to a fixed uniform density in $(0,1)$. An approximate way to do this is to estimate the local variance of the original first, and then to divide (*normalize*) the central coefficient by the square root of that estimate. The resulting normalized coefficient will be approximately Gaussian with unit variance. Then we apply a fixed erf function to make it uniformly distributed. Note that the whole process can be seen as applying a fixed non-linear mapping (which only depends on the noise level) which would transform a GSM vector into an approximately uniformly distributed coefficient.

It is assumed that the hidden scale factor is not really constant for every neighborhood, but that it is a random field which fluctuates smoothly instead [15], so it can be approximated as constant for small neighborhoods. However, because of its (gentil) change across the neighborhood, we will give less weight in the local variance estimation to the outer coefficients than to the coefficients near the center Therefore, instead of using the classical ML estimator $\hat{\sigma_x^2} = \lfloor \mathbf{y}^T\mathbf{y}/N - \sigma_w^2 \rfloor_+$ we will use $\hat{\sigma_x^2} = \lfloor \mathbf{x}^T \mathbf{D_h} \mathbf{x} - \sigma_w^2 \rfloor_+$, where $\mathbf{D_h}$ is a diagonal matrix with its diagonal made by a smoothly varying kernel $\mathbf{h}$, whose samples add up to one. Then, the non-linear transform is:

$$\begin{aligned} f(y;\mathscr{N}(y)) &= erf\,(y/\hat{\sigma}_x(\mathscr{N}(y))) \\ &= erf\left(\frac{y}{\sqrt{\lfloor \sum_{y_i\in\mathscr{N}(y)} h_i y_i^2 - \sigma_w^2 \rfloor_+}}\right) \end{aligned} \quad (10)$$

Instead of using a global inversion model of the coefficients (as in [10]), we use the simplification of keeping the local estimated variance as an intermediate step and do the scalar inversion of $\mathbb{E}\{f(x;\mathscr{N}(x))|y\}$, as explained in section 4.1. Finally, it is convenient to express Eq. 9 in terms of the gain $k = \sigma_x/\sigma_p$ of the function $f(x)$ w.r.t. the *infomax* function ($\sigma_p = \sigma_x$):

$$\hat{x}_{PBLS} = \frac{\hat{\sigma_x^2}(\mathscr{N}(y))}{\sqrt{(\hat{\sigma_x^2}(\mathscr{N}(y)) + \sigma_w^2)(\hat{\sigma_x^2}(\mathscr{N}(y)) + \sigma_w^2(1+k^2))}}y. \quad (11)$$

where we have also substituted the estimated local variance of the uncorrupted signal, $\hat{\sigma}_x{}^2(\mathscr{N}(y))$ by σ_x^2 in Eq. 9.

5. RESULTS AND DISCUSSION

For the results shown here we have used the full steerable pyramid [21, 15] with 5 scales and three orientations, and the non-linearity described in Section 4.2. We have considered only spatial neighbors within the same subband around each coefficient. For estimating the local variance we have used an isotropic Gaussian kernel $\mathbf{h}$ with $\sigma_h = 2$. This relatively big size (hand-optimized) provokes that only rarely the denominator of Eq. 10 vanishes, thus avoiding saturation values in $f(y)$. After estimating all the coefficients, the image is reconstructed by inverting the pyramid.

We have carried out three experiments: (1) using $k = 0$ in Eq. 11 (classical local adaptive Wiener denoising); (2) using $k = 1$ (*infomax* local PBLS solution); and (3) using $k = 2$ ("over-sensitive" PBLS). We have used four noise levels (σ_w = 5, 15, 25, 50) and three standard gray-level 512×512 images: *Lena*, *Boats* and *Barbara*. The solution proposed here is computationally efficient. Processing a 512×512 image takes 12 s. with our non-optimized Matlab®implementation using a PIV 2.0 GHz. The numerical results (in Peak Signal to Noise Ratio, defined as $10\log_1 0(255^2/MSE)$, in decibels) are shown in Table 1. These are preliminary results with a very simple model (e.g., contrary to [15], no correlation among samples or posterior density are considered). Thus, we have not aimed here to achieve state-of-the-art performance, but to demonstrate the interest of this approach for future applications.

Some visual results are shown in Fig. 2 (contrast enhanced). We have used white Gaussian noise of $\sigma_w = 25$. From left to right and top to bottom, we show the noisy image, the result using $k = 0$, $k = 1$ and $k = 2$. We can see how the estimation has improved with respect the classical BLS estimation, both visually and in PSNR terms (see table). In particular, we observe that isolated noisy artifacts have decreased significantly their amplitude, whereas the extra shrinkage performed by the estimator has had a relative small effect on image features such as edges, corners, etc. Overall, there is a shift of the estimation towards removing more noisy artifacts, at the price of a slight image feature blurring. It seems clear that this new balance is more visually pleasant than that of the classical approach (up right).

Another point to comment is how the proposed method has increased the quality of the results *also* in MSE terms (see Table 1). We think that this is because the error-cost function in the PBLS estimation plays a similar role as the prior in the standard Bayesian approach. I.e., it provides information about the image statistics, through the adaptation of the error-cost function to these statistics. Furthermore, the *infomax* principle gives the same "representation space" to equiprobable events, so more typical vectors are more likely to come out than rare vectors. Therefore, although not designed to minimize the Euclidean distance of the estimation to the original, in absence of a global prior, the *infomax*-PBLS estimation tends to reduce the error, compared to using a purely quadratic error-cost function.

It is noteworthy that by including an extra-gain ($k = 2$) in the non-linear function we have improved the results in visual terms (less noisy artifacts), and, for high noise, also in MSE terms (see table). Note that here, contrary to other

Figure 2: From left to right and top to bottom: noisy, local Wiener in the pyramid domain ($k = 0$), idem with infomax adaptation ($k = 1$), and using "over-adaptation" ($k = 2$).

works (e.g., [16, 15]) we have not included a prior on the global image. Then, by increasing the sensitivity of the *perceptual* transform, we achieve a similar qualitative effect as by using a prior in the computation of $p(\mathbf{x}|\mathbf{y})$, because both the saturating non-linearity and the prior cause the shrinkage of the signal (as shown in Fig. 1). As a consequence, results improve w.r.t. using the *infomax* non-linearity ($k = 1$). This hypothesis should be tested in the future, by including a global prior in the model and seeing if results still improve using $k > 1$. Also, considering that in the HVS there is internal noise, amplifying relatively more the low amplitude than the high amplitude responses seems a sensible strategy (e.g., Dolby® system), even when a global prior is considered.

k/σ_w	5	15	25	50
Lena				
0	38.14	32.33	29.36	24.96
1	**38.33**	32.78	29.99	25.94
2	38.24	**32.92**	**30.35**	**26.77**
Barbara				
0	37.34	30.60	27.48	23.33
1	**37.53**	**30.86**	**27.80**	23.90
2	37.45	30.66	27.66	**24.15**
Boats				
0	36.79	30.76	27.98	24.00
1	**36.92**	**31.05**	28.40	24.73
2	36.72	30.98	**28.48**	**25.21**

Table 1: Denoising performance expressed as Peak Signal-to-Noise Ratio in dB for the 3 tested models.

REFERENCES

[1] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Pat. Anal. Mach. Intell.* **11**, pp. 674–693, July 1989.

[2] G. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consumer Electronics*, **38**, pp. 18-34, 1992.

[3] P. Teo and D. Heeger, "Perceptual image distortion," in *Proc. of 1st. Int. Conf. on Image Proc.)*, **2**, pp. 982–986, 1994.

[4] Z. Wang, A.C. Bovik, and E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. on Image Proc.*, **13**, no. 4, pp. 600-612, April 2004.

[5] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Computation* **7**(6), pp. 1129–1159, 1995.

[6] H. B. Barlow, "Possible principles underlying the transformation of sensory messages," in *Sensory Communication*, W. A. Rosenblith, ed., pp. 217–234, MIT Press, 1961.

[7] E. Simoncelli and B. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience* **24**, pp. 1193–1216, May 2001.

[8] H. Barlow, "Redundancy Reduction Revisited," *Network: Computation in Neural Systems*, **12**, pp. 241-253, 2001.

[9] J. Malo and J. Gutierrez, "V1 non-linearities emerge from local-to-global non-linear ICA,", *Network: Computation in Neural Systems*, in press.

[10] J. Malo, E.P. Simoncelli, I. Epifanio, and R. Navarro, "Non-linear image representation for efficient perceptual coding," *IEEE Transactions on Image Processing*, **15**, pp. 68–80, 2006.

[11] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature* **381**, pp. 607–609, 1996.

[12] E. P. Simoncelli, "Bayesian denoising of visual images in the wavelet domain," in *Bayesian Inference in Wavelet Based Models*, P. Müller and B. Vidakovic, eds., ch. 18, pp. 291–308, Springer-Verlag, New York, 1999.

[13] 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.* **11**, pp. 545–557, 2002.

[14] L. Şendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Proc.* **50**, pp. 2744–2756, 2002.

[15] 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.* **12**, pp. 1338–1351, 2003.

[16] 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* **6**, pp. 300–303, 1999.

[17] X. Li and M. T. Orchard, "Spatially adaptive image denoising under overcomplete expansion," in *IEEE Int'l Conf on Image Proc*, **3**, pp. 300–303, IEEE, (Vancouver), 2000.

[18] 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*, pp. 37–40, IEEE, (Thessaloniki), 2001.

[19] D. Andrews and C. Mallows, "Scale mixtures of normal distributions," *J. Royal Stat. Soc.* **36**, pp. 99–102, 1974.

[20] 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., **12**, pp. 855–861, MIT Press, 2000.

[21] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. 2nd Int'l Conf on Image Proc*, **III**, pp. 444–447, IEEE, (Washington, DC), 1995.