

# RECOVERING WAVELET RELATIONS USING SVM FOR IMAGE DENOISING

Valero Laparra\*, Juan Gutiérrez†, Gustavo Camps-Valls‡ and Jesús Malo\*

\*Dept. d'Òptica, †Dept. Informàtica, ‡Dept. Eng. Electrònica, Universitat de València  
Dr. Moliner, 50 - 46100 Burjassot, València, Spain

{valero, jgutierrez, gcamps, jmallo}@uv.es

## ABSTRACT

Here we propose an alternative non-explicit way to take into account the relations among wavelet coefficients in natural images for denoising: we use Support Vector Machines (SVM) to *learn* these relations. Since relations among the coefficients are specific to the signal, SVM regularization removes the noise, which does not share this property. Moreover, due to its non-parametric nature, the method can eventually cope with different noise sources. The results show that: (1) the proposed non-parametric method outperforms conventional methods that assume coefficient independence, and (2) its performance is similar to state-of-the-art parametric methods that do explicitly include these relations. Therefore, the proposed machine learning approach can be seen as a more flexible (model-free) alternative to the explicit description of wavelet coefficient relations in Bayesian approaches.

*Index Terms*— SVM, denoising, wavelet, natural images.

## 1. INTRODUCTION

Denoising requires representing the distorted signal in a domain where signal and noise display different enough behavior. In such a representation, noise is removed by imposing the known properties of the signal to the distorted samples. For example, regularization techniques impose smoothness in the spatial domain since noise is typically uncorrelated.

Wavelet representations are better suited to represent higher order relations among image samples. In fact, Independent Component Analysis (ICA), designed to remove higher order statistical relations, gives rise to local frequency (wavelet-like) basis functions when applied to a set of natural images. Wavelet representations are quite convenient in image denoising because natural image samples have a very specific statistical behavior in this domain. On the one hand, smoothness is represented by a strong energy compaction in coarse scales. On the other hand, the combination of smooth regions with local, high contrast features, such as edges, gives rise to sparse activation of wavelet sensors. This leads to very particular, heavy-tailed, marginal PDFs. These basic features were incorporated in the classical wavelet-based image denoising techniques [1–3]. Classical techniques such as

hard and soft thresholding [1] have been generalized by using Bayesian approaches, looking for *Maximum a Posteriori* (MAP) or *Least Squares* (LS) estimators, together with simple marginal models and assuming statistical independence among coefficients [2, 3].

Marginal models in the wavelet domain are not enough for a proper signal characterization. As images do not really come from a linear combination of independent sources (as assumed by linear ICA), relevant relations among coefficients still remain after wavelet transforms [2]. Relations among wavelet coefficients have proven to be a key issue in applications such as image coding [4, 5] or texture synthesis [6]. The use of these relations is in the core of the most recent and successful image denoising approaches [7, 8]. In this case, more complex image models explicitly including the relations among coefficients have to be plugged and fitted into the Bayesian framework to obtain the image estimates.

Unfortunately, all these model-based Bayesian techniques have three common problems: (1) they critically depend on the accuracy of the image model; (2) MAP or LS estimations can only be analytically derived for particular, typically Gaussian noise sources (for different noise sources, the whole technique has to be reformulated and may not be analytically tractable); and (3) image model parameter estimation from the noisy observation is difficult in general. On the contrary, non-parametric approaches can include the above qualitative properties in an indirect way without the restriction of being analytically attached to particular image or noise models. In this way, in principle, their applicability may be more general.

In this work, we propose recovering the natural relations among wavelet coefficients by using non-parametric support vector machines (SVM) on the noisy image. Since these relations are an intrinsic feature of the signal, the SVM regularization removes the noise, which does not share this property. We focus on the ability of the proposed SVM method to take into account these relations with regard to model-based Bayesian methods. Table 1 shows the properties of different previously reported approaches and the proposed SVM method<sup>1</sup>. This classification shows three eventual advantages

<sup>1</sup>Hard and soft thresholding were classified as model-dependent since they can be derived using a very specific combination of noise and image (MAP estimation assuming Gaussian noise, and Generalized Laplacian PDF signal with particular kurtosis [2, 3]), thus being otherwise suboptimal.

**Table 1.** Properties of the considered denoising methods.

Method	Indep. of Image Model	Indep. of Noise Model	Signal Relations
<i>Hard thres. (HT)</i> [1]	×	×	×
<i>Soft thres. (ST)</i> [1]	×	×	×
<i>Bayes.Gen.Lapl.(BL)</i> [2]	×	×	×
<i>Bayes.Gauss. (BG)</i> [3]	×	×	×
<i>Bayes. GSM</i> [7]	×	×	✓
SVM	✓	✓	✓

of the proposed approach: (1) it does not use a particular parametric image model to be fitted; (2) its solution may be found for complex noise sources even without knowing the functional form of the noise PDF, since it can work with just noise histograms; and (3) it is capable to take into account the relations among wavelet coefficients of natural images. Therefore, the proposed machine learning approach can be seen as a more flexible (model-free) alternative to the explicit description of wavelet coefficient relations for image denoising.

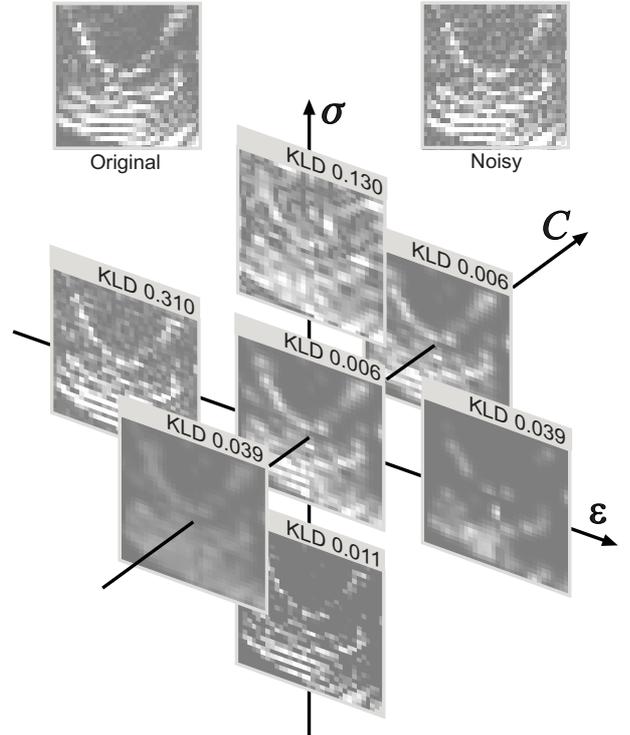
Non-explicit use of dependencies in local frequency domains for denoising was also introduced in [9]. In that case, relations were embedded into a perceptual model used for non-parametric spectrum estimation. Here we pursue the same goal (a model-free technique) with a completely different framework (SVM instead of perceptual information). The idea of using SVM regularization in the wavelet domain for image denoising was introduced in [10], but in this preliminary work, (1) the qualitative effect of the different parameters of the SVM was not analyzed, (2) these parameters were set *ad-hoc*, and more importantly (3) the relevance of the relations among the wavelet coefficients of the signal was not an issue, so the ability of SVM to learn these relations was not assessed nor compared to other methods that do consider them.

The paper is outlined as follows. Section 2 motivates the need for regularization in the wavelet domain, and how SVM can address this issue. Section 3 presents a criterion to select the SVM that better identifies the noise PDF and gives rise to an appropriate (smooth) signal estimation. In Section 4, the performance of the proposed non-parametric method is compared to all parametric Bayesian methods in Table 1. Finally, Section 5 draws some conclusions and further work.

## 2. RESTORING WAVELET RELATIONS WITH SVM

The effect of noise in the wavelet domain is introducing artificial deviations on top of the original signal and hiding the natural relations among the coefficients (see an illustrative example at the top panel of Fig. 1). The regularization ability of SVM [11] can be applied on the noisy signal to remove these artificial deviations while recovering (or imposing) the natural relations among the coefficients.

Throughout this work, the degraded observation,  $\mathbf{i}_d$ , is obtained from an original image,  $\mathbf{i}_o$ , plus an unknown realization of additive noise,  $\mathbf{n}$ , with known PDF or histogram,



**Fig. 1.** *Top:* magnitude of original and noisy wavelet subband patches of ‘Barbara’ image. *Bottom:* effect of SVM parameters on the noisy patch. KL-divergence values between the estimated and the actual noise PDFs are given in each case (see text).

$P(\mathbf{n})$ . A wavelet transform,  $T$ , is applied to the observed image, leading to a set of (noisy) coefficients,  $\mathbf{y} = T \cdot \mathbf{i}_d$ .

Now, given an input-output pair of  $N$ -dimensional vectors  $\{\mathbf{x}_i, y_i\}_{i=1}^N$ , where  $\mathbf{x}_i$  are the wavelet indices and  $y_i$  are the noisy wavelet coefficients, and a non-linear mapping  $\phi$  to a higher dimensional feature space, the SVM computes the weights  $\mathbf{w}$  to obtain the estimation,  $\hat{y}_i = \phi(\mathbf{x}_i)^\top \mathbf{w}$ , minimizing the following *regularized* functional:

$$\|\mathbf{w}\|^2 + C \sum_i \xi_i \quad (1)$$

subject to  $|y_i - \phi(\mathbf{x}_i)^\top \mathbf{w}| \leq \epsilon + \xi_i, \forall i = 1, \dots, N$ , where  $\xi_i$  are the magnitude of the deviations of the estimated signal from the observed noisy data outside the insensitivity zone  $\epsilon$ . Parameter  $C$  tunes the trade-off between fitting the model to the observed noisy data (minimizing the deviations) and keeping model weights  $\|\mathbf{w}\|$  small (enforcing flatness in the feature space). It is worth stressing that parameter  $\epsilon$  accounts for the allowed error, and it is strongly linked to the noise variance [12]. Explicitly working with the non-linearity  $\phi$  is no longer necessary since the whole formulation can be expressed in the form of dot products of the mapping functions called *kernels*,  $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$ . In this case, the estimation is given by  $\hat{\mathbf{y}} = K\alpha$ , where  $\alpha$  is the dual representation of weights  $\mathbf{w}$  [11]. The kernel matrix can be seen

as a similarity matrix among samples (or coefficients), and should reflect the relations among them. Many kernel functions have been proposed in the literature [11]. Here we used Radial Basis Functions (RBF) in the kernel since the relationship among the wavelet coefficients corresponding to spatial neighbors within a subband is local.

According to the above, the regularization behavior of the SVM depends on three parameters: the insensitivity zone,  $\varepsilon$ , the extent of the neighborhood defined by the kernel (the width,  $\sigma$ , in the particular RBF choice), and the penalization factor,  $C$ . The bottom panel of Fig. 1 shows the qualitative effect of SVM estimation as a function of these parameters. Increasing the kernel width,  $\sigma$  (vertical direction), introduces too strong relations among coefficients in such a way that spurious energy appears. Increasing the insensitivity,  $\varepsilon$  (horizontal direction), a sparser solution is obtained, leading to information loss and thus relevant features of the signal are discarded. On the contrary, too small insensitivity gives rise to overfitting, and hence noise is not removed. Small values of the  $C$  parameter gives rise to over-regularized estimations.

According to the above example, dedicated SVMs can certainly discover the underlying structure of the original signal from the noisy observation, but a criterion to choose the appropriate SVM is needed.

Fitting an arbitrary SVM on the noisy data in the transformed domain, gives rise to a particular image estimation,  $\hat{\mathbf{i}} = T^{-1} \hat{\mathbf{y}}$ . The difference between this estimation and the noisy input is the *estimated noise*,  $\mathbf{n}_e = \mathbf{i}_d - \hat{\mathbf{i}}$ . In general,  $\mathbf{n}_e \neq \mathbf{n}$ , however, one can always find the SVM that maximizes the similarity between the estimated and actual noise probability functions (or histograms),  $P(\mathbf{n}_e)$  and  $P(\mathbf{n})$ , respectively. Our proposal considers minimizing the Kullback-Leibler divergence (KLD) between the joint PDFs to enforce similarity up to the higher (computationally estimable) order. The underlying idea is that the SVM that enforces the similarity between the estimated noise and the actual noise is the one that best captures the relevant features of the true signal.

The proposed scheme is depicted in Fig. 2. The noisy image is transformed to a wavelet domain where intraband coefficients are grouped into  $N$ -dimensional vectors because, in images, the mutual information is typically larger among intraband than interband neighbors. Starting from an arbitrary set of SVM parameters,  $\theta = \{\varepsilon, \sigma, C\}$ , the feedback loop in Fig. 2 selects the SVM with parameters

$$\theta^* = \arg \min_{\theta} \left\{ \text{KLD}[P(\mathbf{n}_e(\theta)), P(\mathbf{n})] \right\}, \quad (2)$$

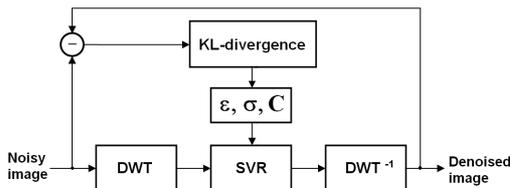


Fig. 2. The proposed denoising algorithm.

Table 2. Similarities/errors of the estimations at  $\sigma_n^2 = 400$ .

	'Barbara'		'Boats'		'Lena'	
	SSIM	RMSE	SSIM	RMSE	SSIM	RMSE
HT	0.74	13.69	0.72	13.00	0.73	11.84
ST	0.74	11.96	0.72	11.75	0.71	10.78
BG	0.75	11.99	0.73	11.64	0.73	10.68
BL	0.75	11.45	0.77	10.27	0.72	12.36
GSM	0.87	8.72	0.84	8.47	0.85	7.99
SVM	0.85	9.76	0.81	11.01	0.81	9.95

leading to an estimation with appropriate relations among coefficients. KLD values in Fig. 1 illustrate the usefulness of the above procedure: the minimum divergence point (central patch) is also the best estimation of the original patch.

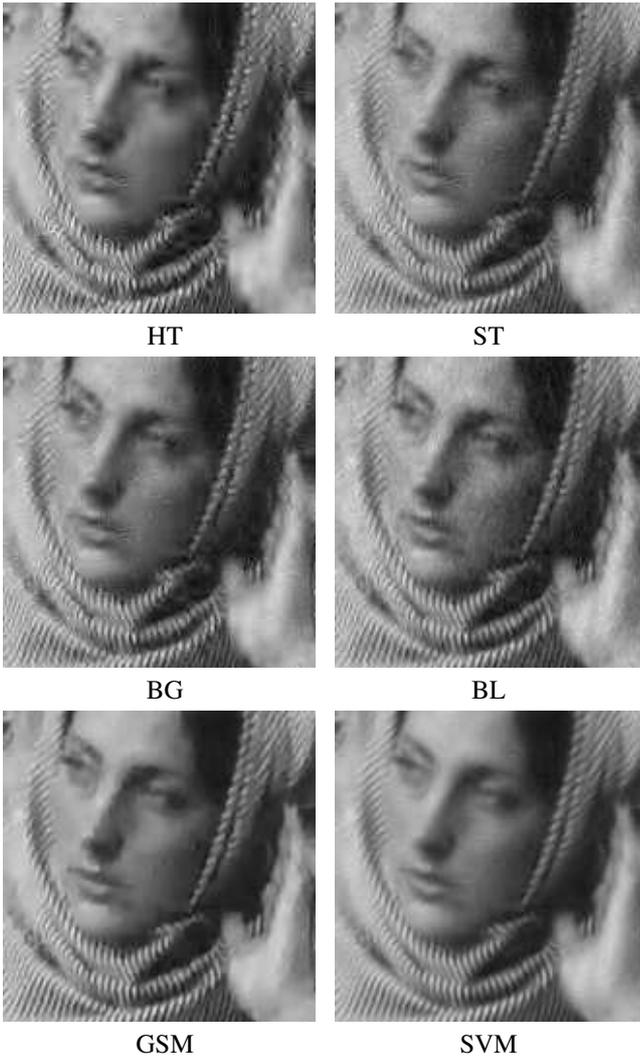
All the analyzed Bayesian algorithms in Table 1 rely on a particular analytical noise PDF (typically Gaussian of known covariance). The advantage of our method is that KLD can be computed regardless of the complexity this PDF or even in the case that only the noise histogram is available.

### 3. EXPERIMENTAL RESULTS

We used an 8-orientation, 4-scale, steerable wavelet transform,  $T$  [13]. Subbands are decomposed into 256-dimensional vectors  $\mathbf{y}$ . In eq. (2) we restricted ourselves to 2-dimensional joint histograms due to the lack of noise samples yet enforcing decorrelation in the estimated noise. SVM optimization was carried out through exhaustive search. A scale-dependent  $\varepsilon_i$  profile [14] was used according to the noise variance in the selected wavelet domain. This profile was scaled by a factor varied in the range [0.4, 2.6], which was chosen according to [12]. The penalization parameter  $C$  was varied in the range [10,  $10^4$ ], according to the prescription in [15] related to the signal variance. The kernel width  $\sigma$  was varied in the range [0.01, 3.5] in spatial coefficient units, to cover a wide range of possible intrascale signal relations.

Our algorithm was compared to all methods in Table 1 using different standard  $256 \times 256$  images with different levels of degradation (Gaussian noise with  $\sigma_n^2 = \{100, 200, 400\}$ ). Table 2 shows representative distortion results measured using the standard (yet not perceptually meaningful) RMSE, and the perceptually meaningful Structural SIMilarity (SSIM) index [16]. Results show that our algorithm performs better than the methods that neglect signal relations (HT, ST, BG and BL), and obtains similar (yet slightly lower) numerical results than the one which incorporates them (GSM). The method performs consistently through all images. Besides, the same was observed for all the considered noise variances, thus suggesting that the guiding criterion is robust.

Figure 3 shows representative results in the challenging situation of  $\sigma_n^2 = 400$ . It can be noted that thresholding methods (HT, ST) and Bayesian generalizations not including relations in the model (BG, BL) show poor performance, producing images either grained or corrupted by too salient wavelet



**Fig. 3.** Visual results for ‘Barbara’,  $\sigma_n^2 = 400$ .

functions. Even though SVM yields slightly lower numerical scores than GSM, global visual performance is comparable. Nevertheless, our method reveals better in some aspects: GSM produces some residual patches, which are not present in our solution, and also high and low frequency details (e.g., see Barbara’s scarf and nose) are better preserved.

#### 4. CONCLUSIONS

In this work, we proposed a novel image denoising procedure based on the use of SVM in a steerable wavelet domain. A KLD-based criterion was proposed to automatically select the SVM that best recovers the relevant wavelet coefficient relations of the true signal. The results show that the performance of the proposed non-parametric method (1) is better than conventional methods that assume coefficient independence, and (2) is similar to state-of-the-art methods that do explicitly in-

clude these relations. Therefore, the proposed SVM approach can be seen as a more flexible (model-free) alternative to the explicit description of coefficient relations. Future work is tied to the incorporation of new information in the kernel, and to test the flexibility of the method in restoring images corrupted by more general noise sources.

#### 5. REFERENCES

- [1] David L. Donoho and Iain M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *J. Am. Stat. Assoc.*, vol. 90, pp. 1200–1224, 1995.
- [2] E. Simoncelli, “Bayesian denoising of visual images in the wavelet domain,” in *Bayesian Inference in Wavelet Based Models*, pp. 291–308. Springer-Verlag, New York, 1999.
- [3] M. Figueiredo and R. Nowak, “Wavelet-based image estimation: an empirical Bayes approach using Jeffrey’s noninformative prior,” *IEEE TIP*, vol. 10, no. 9, pp. 1322–1331, 2001.
- [4] J. Malo, I. Epifanio, R. Navarro, and E. Simoncelli, “Non-linear image representation for efficient perceptual coding,” *IEEE Trans. Im. Proc.*, vol. 15, no. 1, pp. 68–80, 2006.
- [5] G. Camps-Valls, J. Gutiérrez, G. Gómez, and J. Malo, “On the suitable domain for SVM training in image coding,” *To be published in: J. Mach. Learn. Res.*, vol. 9, 2008.
- [6] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *Int. J. Comp. Vis.*, vol. 40, no. 1, pp. 49–71, 2000.
- [7] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, “Image denoising using a scale mixture of Gaussians in the wavelet domain,” *IEEE TIP*, vol. 12, no. 11, pp. 1338–1351, 2003.
- [8] L. Simeï and E. Simoncelli, “Statistical modeling of images with fields of GSMs,” in *Proc. NIPS06*. May 2007, MIT Press.
- [9] J. Gutiérrez, F. Ferri, and J. Malo, “Regularization operators for natural images based on nonlinear perception models,” *IEEE Trans. Im. Proc.*, vol. 15, no. 1, pp. 189–200, 2006.
- [10] H. Cheng, J.W. Tian, J. Liu, and Q.Z. Yu, “Wavelet domain image denoising via SVR,” *Electr. Lett.*, vol. 40, 2004.
- [11] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Stats. and Comp.*, vol. 14, pp. 199–222, 2004.
- [12] J. T. Kwok and I. W. Tsang, “Linear dependency between  $\varepsilon$  and the input noise in  $\varepsilon$ -Support Vector Regression,” *IEEE Trans. Neur. Nets.*, pp. 544–553, May 2003.
- [13] E. Simoncelli and W. Freeman, “The steerable pyramid: A flexible architecture for multi-scale derivative computation,” in *Proc 2nd IEEE Int’l Conf. Im. Proc.*, 1995.
- [14] G. Gomez, G. Camps-Valls, J. Gutierrez, and J. Malo, “Perceptual adaptive insensitivity for support vector machine image coding,” *IEEE TNN*, vol. 16, no. 6, pp. 1574–1581, 2005.
- [15] V. Cherkassky, “Practical selection of SVM parameters and noise estimation for SVM regression,” *Neur. Nets.*, vol. 17, no. 1, pp. 113–126, 2004.
- [16] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Im. Proc.*, vol. 13, no. 4, pp. 600–612, 2004.