Perceptual Adaptive Insensitivity for Support Vector Machine Image Coding

Gabriel Gómez-Pérez, Gustavo Camps-Valls, *Member, IEEE*,
Juan Gutiérrez, and Jesús Malo.

Abstract

Support Vector Machine (SVM) learning has been recently proposed for image compression in the frequency domain using a constant ε-insensitivity zone by Robinson and Kecman [1]. However, according to the statistical properties of natural images and the properties of human perception, a constant insensitivity makes sense in the spatial domain but it is certainly not a good option in a frequency domain. In fact, in their approach, they made a fixed low-pass assumption as the number of DCT coefficients to be used in the training was limited. This paper extends the work of Robinson and Kecman by proposing the use of adaptive insensitivity SVMs [2] for image coding using an appropriate distortion criterion [3], [4] based on a simple visual cortex model. Training the SVM by using an accurate perception model avoids any *a priori* assumption and improves the rate-distortion performance of the original approach.

Index Terms

Support Vector Machine, Adaptive Insensitivity, Image Coding, DCT, Perceptual Metric, Maximum Perceptual Error.

Gabriel Gómez-Pérez and Gustavo Camps-Valls are with Dept. Enginyeria Electrònica, Universitat de València, Spain (gabriel.gomez@uv.es, gustavo.camps@uv.es). Juan Gutiérrez is with Dept. Informàtica, Universitat de València, Spain (juan.gutierrez@uv.es). Jesús Malo is with Dept. d'Òptica, Universitat de València, Spain (jesus.malo@uv.es).
Address for correspondence: Gustavo Camps-Valls, Escola Tècnica Superior d'Enginyeria, Dept. Enginyeria Electrònica, Universitat de València, C/Dr. Moliner 50, 46100 Burjassot, València, Spain.

This work has been partially supported by the CICYT project TIC2003-01504 and the GV project Grupos04-08.

I. INTRODUCTION

A recent approach to machine learning problems is the Support Vector Machine (SVM) [5]. The Support Vector Regressor (SVR) [6] is its implementation for function approximation. Several applications of SVM have appeared in the context of image processing, such as face recognition [7], image classification [8], texture segmentation [9], and image fusion [10]. The use of SVMs for image compression was originally presented in [11], where the authors used the SVR to learn the gray levels in the image. However, the statistical properties of the natural images make the Discrete Cosine Transform (DCT) suitable for image representation [12], improving the performance of the SVM learning [1]. According to these results [1], the ability of SVMs to model DCT-transformed image representations with a small set of parameters make them a promising alternative to classical transform coding techniques based on quantization [13], [14].

However, the proposed SVM schemes for image compression have always used a fixed accuracy level (ε-insensitivity) *per* sample [1], [11]. A constant insensitivity zone makes sense in the spatial domain because of the approximate stationary behavior of the luminance samples of natural images. Moreover, the perceptual relevance of pixels is also approximately constant across the spatial domain. However, these facts are no longer true in a frequency domain: the statistics of frequency coefficients of natural images is highly non-stationary and their perceptual relevance is highly uneven [15]. The method proposed by Robinson and Kecman [1] limited the number of DCT coefficients to a fixed number. This approach can affect the reconstructed image by blurring some details in the image, such as sharp edges or high frequency components. This suggests that their results can be improved if the SVM learning in the DCT domain is modulated by a perceptually-based frequency-dependent insensitivity zone.

In order to obtain a good subjective performance in image coding applications, it is important to restrict the Maximum Perceptual Error (MPE) in each DCT coefficient [3], [4], [15], [16]. In this work, we propose an SVM with adaptive insensitivity zone [2] for image coding, which is based on an appropriate Human Visual System (HVS) model. Therefore, using perception models to design the adaptive insensitivity gives rise to SVM coders which are optimal under the MPE criterion, and there is no need to make any *ad-hoc* (low-pass) assumption in the SVM training.

The structure of the paper is as follows. Section II reviews the adaptive SVM formulation and how it can be used in DCT modeling schemes. Section III motivates the perceptual weighting in the DCT domain and analyzes the use of the MPE criterion in the SVM coding scheme. Section IV shows results of our proposal on benchmark images. Section V ends this paper with some conclusions and further work.

II. ADAPTIVE INSENSITIVITY IN THE SUPPORT VECTOR REGRESSOR

The standard formulation of the SVR model is stated as follows. Given a labeled training data set $\{(\mathbf{x}_i, y_i),\ i = 1, ..., n\}$, where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, and a nonlinear mapping to a higher dimensional space $\boldsymbol{\phi} : \mathbb{R}^d \rightarrow \mathbb{R}^H$ where $d \leq H$, solve

$$\min_{\mathbf{w},\xi_i,\xi_i^*,b} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i (\xi_i + \xi_i^*) \right\} \tag{1}$$

subject to:

$$y_i - \boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} - b \leq \varepsilon + \xi_i \qquad \forall i = 1, \ldots, n \tag{2}$$

$$\boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b - y_i \leq \varepsilon + \xi_i^* \qquad \forall i = 1, \ldots, n \tag{3}$$

$$\xi_i, \xi_i^* \geq 0 \qquad \forall i = 1, \ldots, n \tag{4}$$

where $\xi_i^{(*)}$ and C are, respectively, positive slack variables to deal with training samples with a prediction error larger than ε ($\varepsilon > 0$) and the penalization applied to these. The usual procedure for solving SVRs introduces the linear restrictions (2)-(4) into Eq. (1) by means of Lagrange multipliers $\alpha_i^{(*)}$, computes the Karush-Kuhn-Tucker conditions, and solves the Wolfe's dual problem using quadratic programming (QP) procedures [5], [17].

The regression estimate for a given input vector $\mathbf{x}$ then takes the form

$$\hat{y} = f(\mathbf{x}) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \tag{5}$$

where the inner product $\boldsymbol{\phi}(\mathbf{x}_i)^T \cdot \boldsymbol{\phi}(\mathbf{x})$ is represented with a kernel matrix $K(\mathbf{x}_i, \mathbf{x})$. Note that only samples with non-zero Lagrange multipliers $\alpha_i^{(*)}$ count in the solution and are called *support vectors*. The immediate advantage of the method is that good approximating functions can be obtained with a (relatively) small set of support vectors, leading to the concept of *sparsity* and, in turn, to the idea of inherent compression.

However, the main problem when considering this solution is that we assume that each sample contains *a priori* the same relevance to the modelling, which in general is not true. This can be easily alleviated by using a different penalization factor for each training sample i according to a certain *confidence function* c_i on the samples. This idea can be also extended by using different insensitivity zone ε for each sample. In this work, we use the profiled SVR approach [2], which relaxes or tightens the ε-insensitive region depending on each training sample. Now, the objective function becomes [5]:

$$\min_{\mathbf{w},\xi_i,\xi_i^*,b}\left\{\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i c_i(\xi_i+\xi_i^*)\right\} \tag{6}$$

and restrictions over slack variables become sample-dependent:

$$y_i - \boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} - b \leq \frac{\varepsilon}{c_i} + \xi_i \qquad \forall i = 1,\ldots,n \tag{7}$$

$$\boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b - y_i \leq \frac{\varepsilon}{c_i} + \xi_i^* \qquad \forall i = 1,\ldots,n \tag{8}$$

$$\xi_i, \xi_i^* \geq 0 \qquad \forall i = 1,\ldots,n \tag{9}$$

Therefore, now each sample has its own insensitivity error $\varepsilon_i = \varepsilon/c_i$, which intuitively means that different samples hold different confidence intervals. By including linear restrictions (7)-(9) in the corresponding functional (6), we can follow as in the standard case, which once again constitutes a QP problem.

In the SVR image coding procedure [1], the whole image is first divided in blocks, and then a 2D DCT-transform is applied to each one of them. Then, dedicated SVR models are trained in the frequency domain for each block and the obtained weights are quantized. Therefore, the signal is described by the Lagrange multipliers of the support vectors needed to keep the regression error below the thresholds ε_i. Increasing the thresholds, ε_i, reduces the number of required support vectors, thus reducing the entropy of the encoded image and increasing the distortion. The key point here is choosing ε_i according to a meaningful criterion for the application.

In [2], [18], we designed profiles for the variation of C and ε as a function of the sample in complex pharmacokinetic problems. In [19], profiles were defined in terms of clusters rather than fixed *a priori*. In this paper, we will define the ε-insensitive zone to restrict the Maximum Perceptual Error (MPE) [3], [4], [15], [16] in each coefficient of the DCT. This profile will vary the ε-insensitive region as a function of the frequency in the DCT domain.

III. Maximum Perceptual Error for adaptive insensitivity

The core of the transform coding idea is that the relevance of the coefficients in the DCT-transformed domain is highly uneven. This is because while some coefficients have a big contribution to the distortion, others can be strongly modified without significant loss. In the transform and quantization paradigm [14], the hierarchy of coefficients has led to uneven bit allocation schemes (and non-uniform 1D quantizers for each coefficient) [3], [4], [15], [16]. This implies that the maximum distortion introduced in each coefficient depends on both its frequency and its amplitude. These ideas can be incorporated into the SVM paradigm by considering that the maximum distortion is given by the insensitivity parameter ε. Therefore, the distortion criteria used to design the variable quantizer step in each coefficient could be applied to design an adaptive insensitivity zone in the SVM case.

Classical quantizer design is founded on MSE minimization and gives rise to variable quantization steps based on the variance of the coefficients and their particular probability density function [14]. However, as the coded image has to be judged by a human observer, the criterion should include the sensitivity of the human viewer. In that sense, the introduction of a perceptual metric in average error criteria does not solve the problem because average perceptual error minimization does not imply that every error is below (or proportional) to the perceptual discrimination thresholds. In fact, it has been shown that keeping the distortion proportional to the visibility thresholds (restricting the MPE of each coefficient) leads to better subjective results than minimizing the average perceptual error [3], [4], [15], [16]. Therefore, the bottom line to design the adaptive insensitivity zone of the SVM, which restricts the maximum error in each coefficient, is drawn from the MPE criterion in each coefficient for each particular image region.

In our case, we have to compute the human visual insensitivity for every DCT coefficient from the corresponding slope of an appropriate vision response model. Current models of human visual cortex assume that each region, A, of the input image around some spatial position, s, undergoes a two-stage transform [20], [21]:

$$A \xrightarrow{T} y \xrightarrow{R} r \tag{10}$$

where T is a linear transform in which the input is analyzed by a set of unit-norm oriented local-frequency sensors (V1 neurons) with receptive fields qualitatively similar to the block-DCT basis functions [22]:

$$y_i = \sum_j T_{ij} \cdot A_j \tag{11}$$

and R is a transduction function that represents the gain of each particular sensor, T_i, and maps the linear transform representation into a perceptually Euclidean response representation [21]. The Euclidean nature of the response representation implies that the linear transform representation, y, *is not* Euclidean [23].

In this way, a small distortion in the transform representation, Δy, induces a distortion that can be approximated by using the Jacobian of the transduction function:

$$r + \Delta r \simeq R(y) + \nabla R(y) \cdot \Delta y \tag{12}$$

Then, the maximum perceptual distortion for that spatial region is given by

$$\text{MPE}_s = \|\Delta r\|_\infty = \max(\nabla R(y) \cdot \Delta y) \tag{13}$$

The global perceived distortion in an image with n spatial regions will be a particular spatial pooling (β-norm) of these n local distortions from each local (block) response representation:

$$\text{MPE} = \|(\text{MPE}_1, \cdots, \text{MPE}_n)\|_\beta = \left(\sum_s \text{MPE}_s^\beta \right)^{1/\beta} \tag{14}$$

where β is the summation exponent in this spatial pooling. The most accurate gain control models of V1 sensors include non-linearities with interactions between the outputs of the linear sensors [20], [21], thus giving rise to a non-diagonal input-dependent Jacobian [23]. Using such models would not be easy to derive a bound, ε_i, for the distortion in each coefficient, Δy_i, from Eq. (13). However, if we restrict ourselves to the most simple model in which each sensor has a constant linear gain given by the *Contrast Sensitivity Function (CSF)* [24]:

$$\Delta r_i = \text{CSF}_i \cdot \Delta y_i, \tag{15}$$

the Jacobian is a diagonal matrix with $\nabla R(y)_{ii} = \text{CSF}_i$. According to this, in order to keep the perceptual error below some arbitrary threshold, $\text{MPE}_s = \tau$, every distortion, Δy_i, has to be:

$$\Delta y_i \leq \tau \cdot \text{CSF}_i^{-1} \tag{16}$$

Therefore, the insensitivity region for each coefficient y_i should be given by the CSF:

$$\varepsilon_i = \tau \cdot \text{CSF}_i^{-1} \tag{17}$$

Figure 1 shows the CSF, i.e. the relative slope for each sensor (or basis function) of the DCT representation, which is expressed in cycles/degree. The behavior of the visual system in the frequency domain (e.g. the CSF) is commonly defined in physically meaningful units such as cycles/degree or samples/degree. These units refer to the number of discrete samples per angle subtended by the image at a given viewing distance. The frequency meaning of the DCT coefficients is given by the selected sampling frequency (or equivalently by the size and viewing distance).

The discrimination ability of a sensor (its insensitivity ε) can be obtained from the slope of its response curve. Figure 2 shows that the bigger the slope, the smaller the insensitivity: different slopes in the response of each sensor imply different insensitivities, and hence different bounds on Δy_i for the same perceptual error $\text{MPE}_s = \tau$.

Using insensitivity values according to Eq. (17) is optimal in the MPE sense because it ensures that the MPE_s is below the selected threshold, τ, for every region, s, thus minimizing the global MPE.

IV. Results and Discussion

The general encoding procedure proposed by Robinson and Kecman [1] consists of learning the DCT representation of each block of the image to obtain a set of support vectors and their corresponding Lagrange multipliers. These weights are then uniformly quantized. The number of selected support vectors and thus the entropy of the encoded signal is controlled by a factor applied to the ε-insensitivity zone (the parameter τ in Eq. (17)). Tailoring different ε profiles will produce critically different support vector distributions in the frequency domain and hence different error distributions in this domain. Therefore, different ε profiles lead to results of quite different perceptual quality.

In this section, we show the benefits of the proposed MPE optimal profile (CSF-SVR approach, Eq. (17)) by comparing its results with a generic uniform tube (ε-SVR approach), and with the method proposed by Robinson and Kecman [1] (RKi-1 approach). We compare these three different SVM training strategies in terms of (a) the distribution of support vectors, and (b) the effect that these distributions have in the compression performance. Following the same approach of [1], we used the RBF kernel, trained the SVR models without the bias term b, and modeled the absolute value of the DCT coefficients. For the sake of a fair comparison, all the free parameters (ε-insensitivity, penalization parameter C, Gaussian width of the RBF kernel, and uniform quantization level of the weigths) were optimized for all the considered models. The value of τ in (17) was tuned iteratively to produce a given compression ratio and depends on the image. Note that high values of τ increase the width of the ε tube, which in turn produce lower number of support vectors and consequently yield higher compression ratios.

A. Distribution of support vectors

Figure 3 shows a representative example of the distribution of the selected SVs by the three models considered in this work. These distributions reflect how the selection of a particular insensitivity profile modifies the learning behavior of the SVMs.

Using a straightforward constant ε for all coefficients (ε-SVR approach) concentrates more support vectors in the low frequency region because the variance of these DCT coefficients in natural images is higher [12], [15]. However, it still yields a relatively high number of support vectors in the high-frequency region. This is inefficient because of the low subjective relevance of that region (see Fig. 1). Considering these vectors will not significantly reduce the (perceptual) reconstruction error while it increases the entropy of the encoded signal.

The RKi-1 approach [1] uses a constant ε but the authors solve the above problem by neglecting the high-frequency coefficients in training the SVM for each block[1]. This is equivalent to the use of an arbitrarily large insensitivity for the high-frequency region. As a result, this approach relatively allocates more support vectors in the low/medium frequency regions. As the authors suggest, this modification of the straightforward uniform approach is qualitatively based in the basic low-pass behavior of human vision. However, such a crude approximation (that implies no control of the distortion in the high-frequency region) can introduce annoying errors in blocks with sharp edges.

The proposed algorithm (CSF-SVR approach) uses a variable ε according to Eq. (17). Taking into account the perception facts reviewed in Section III, the acceptable distortion in the low/medium-frequency region is smaller than in the high-frequency region, giving rise to a (natural) concentration of support vectors in the low/medium frequency region. Note that this

[1]If a (reasonable) sampling frequency of 64 cycles/degree is assumed, the cut-off value recommended in [1] is around 20 cycles/deg.

concentration is even bigger than in the RKi-1 approach. However, the proposed algorithm does not neglect any coefficient in the learning process. This strategy naturally reduces the number of allocated support vectors in the high-frequency region with regard to the straightforward uniform approach, but it does not prevent selecting some of them when it is necessary to keep the error below the selected threshold, which may be relevant in edge blocks.

B. Compression performance

Exhaustive compression experiments using several standard images (Lena, Barbara, Boats, Peppers and Cameraman) were conducted using the different SVM training strategies at different compression rates in the range [0.05, 0.5] bits/pixel (bpp), i.e. 160:1 to 16:1 compression ratios, respectively. In all cases, the images were analyzed using 16$\times$16 block-DCT, assuming a sampling frequency of 64 cycles/degree. We also include results using the standard JPEG as a baseline method for reference purposes[2].

Given the limitations of the available (subjective) distortion metrics [21], [25]–[27], the more reliable evaluation of the subjective performance of the considered methods is the direct visual inspection of the decoded images. However, it is also usual to describe the compression performance using rate-distortion curves. In these curves, the volume of the encoded image (measured, for instance, by its entropy in bits/pixel) is compared to an appropriate distortion measure. The best algorithm is the one that achieves the lowest distortion for a range of bit rates. In this case, the distortion measure should be meaningful for the application, i.e. it should represent the subjective quality of the reconstructed image.

In this section, we analyze the performance of the algorithms through rate-distortion curves using two different distortion measures: the standard MSE$^{1/2}$ and the MPE of Eq. (14) with $\beta = 2$ and using the CSF model for ∇R. Results are shown in Fig. 4. According to the standard MSE point of view, the performance of the SVM algorithms is basically the same (see Fig. 4(a)), improving the results of JPEG as previously reported in [1]. However, we can observe a substantial gain in MPE of the CSF-SVR model when looking at Fig. 4(b). As expected from the discussion in Section III, the proposed scheme is optimal under the MPE criterion (and the CSF model) and, of course, it is suboptimal in the MSE (or PSNR) sense. In fact, by taking into

[2]We used the JPEG implementation by Lagendijk, which is available at `http://www-ict.ewi.tudelft.nl`.

account the visual results presented in Fig. 5, it is clear that the MSE results are not useful to represent the quality of the methods, as extensively reported elsewhere [21], [25]–[27]. These results not only confirm the theoretical and practical validity of incorporating the CSF into the SVM methodology, but also the meaningfulness of the MPE distortion measure [3], [15], [16], [23]. The visual inspection of the results (Fig. 5) confirm that the *numerical* gain in MPE shown in Fig. 4(b) is also *perceptually significant*.

The visual effect of the different distribution of the support vectors due to the different insensitivity profiles is clear in Fig. 5. First, it is obvious that the perceptually-based training leads to better overall subjective results: the annoying blocking artifacts of the ε-SVR and RKi-1 approaches are highly reduced in the proposed approach, giving rise to smoother, and perceptually more acceptable, images. Second, the blocking artifacts in ε-SVR and RKi-1 approaches may come from different reasons. On the one hand, the uniform ε-SVR wastes (relatively) too many support vectors (and bits) in the high-frequency region in such a way that noticeable errors in the low-frequency components (related to the average luminance in each block) are produced (see the face of Barbara). However, note that due to the allocation of more vectors in the high-frequency region, it is the method that better reproduces details such as the high-frequency strips in the Barbara clothes. On the other hand, neglecting the high-frequency coefficients in the training (RKi-1 approach) does reduce the blocking a little bit, but it cannot cope with high contrast edges that also produces a lot of energy in the high frequency region (for instance, Lena's cheek on the dark hair background).

An example of the performance of RKi-1 and CSF-SVR at high compression ratios (from 64:1 to 125:1) is illustrated in Fig. 6 and Table I. Both the numerical and visual results show the same trend observed in Fig. 5. Specifically, the proposed method reduces the blocking effect due to a better perceptually-based distribution of support vectors. The reduction in MPE distortion in Table I is confirmed by the appearance of the CSF-SVR results in Fig. 6.

V. Final remarks

In this work, we have tailored an ε-insensitivity function in the SVR model for image coding, which is optimal under the MPE principle. This approach has been motivated by the fact that, in the DCT-transformed domain, the use of a fixed ε value is not consistent with the statistical and perceptual properties of natural images. This approach has revealed to be more efficient than the original SVR-based coding schemes in terms of perceptually meaningful rate-distortion measure and visual inspection, precluding *ad hoc* assumptions in the training algorithm.

An accurate consideration of a perceptually profiled SVR training has improved the results. This fact suggests that further improvement could be achieved by including more sophisticated non-linear perceptual models [4] in support vector coding schemes.

REFERENCES

[1] J. Robinson and V. Kecman, "Combining Support Vector Machine learning with the discrete cosine transform in image compression," *IEEE Transactions Neural Networks*, vol. 14, no. 4, pp. 950–958, July 2003.

[2] G. Camps-Valls, E. Soria-Olivas, J. Pérez-Ruixo, A. Artés-Rodríguez, F. Pérez-Cruz, and A. Figueiras-Vidal, "A profile-dependent kernel-based regression for cyclosporine concentration prediction," in *Neural Information Processing Systems (NIPS) – Workshop on New Directions in Kernel-Based Learning Methods*, Vancouver, British Columbia, Canada, December 2001, Available at http://www.uv.es/∼gcamps.

[3] J. Malo, J. Gutiérrez, I. Epifanio, F. Ferri, and J. M. Artigas, "Perceptual feed-back in multigrid motion estimation using an improved DCT quantization," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1411–1427, October 2001.

[4] J. Malo, I. Epifanio, R. Navarro, and E. Simoncelli, "Non-linear image representation for efficient perceptual coding," *IEEE Transactions on Image Processing*, 2005, Accepted for publication (in press). Available at http://www.uv.es/vista/vistavalencia/.

[5] Vladimir N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, 1998.

[6] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199–222, 2004.

[7] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, Jan 2003.

[8] O. Chapelle, P. Haffner, and V. Vapnik, "SVMs for histogram-based image classification," *IEEE Transactions on Neural Networks*, vol. 9, no. 5, pp. 1055–1064, 1999.

[9] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Supervised texture segmentation using support vector machines," *IEE Electronics Letters*, vol. 35, no. 22, pp. 1935–1936, 1999.

[10] S. Li, J.T.-Y. Kwok, I.W.-H. Tsang, and Y. Wang, "Fusing images with different focuses using support vector machines," *IEEE Transactions on Neural Networks*, vol. 15, no. 6, Nov 2004.

[11] J. Robinson and V. Kecman, "The use of Support Vector Machines in image compression," in *Proceedings of the International Conference on Engineering Intelligence Systems, EIS'2000*, University of Paisley, Scotland, U.K., June 2000, vol. 36, pp. 93–96.

[12] R.J. Clarke, "Relation between the Karhunen-Loeve transform and cosine transforms," *Proceedings IEE, Pt. F*, vol. 128, no. 6, pp. 359–360, 1981.

[13] G.K. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp. 31–43, April 1991.

[14] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Press, Boston, 1992.

[15] J. Malo, F. Ferri, J. Albert, J. Soret, and J. M. Artigas, "The role of perceptual contrast non-linearities in image transform coding," *Image & Vision Computing*, vol. 18, no. 3, pp. 233–246, February 2000.

[16] J. Malo, F. Ferri, J. Albert, and J. Soret, “Comparison of perceptually uniform quantisation with average error minimisation in image transform coding,” *IEE Electronics Letters*, vol. 35, no. 13, pp. 1067–1068, June 1999.

[17] B. Schölkopf, P. L. Bartlett, A. Smola, and R. Williamson, “Shrinking the tube: a new support vector regression algorithm,” in *Advances in Neural Information Processing Systems 11*, M. S. Kearns, S. A. Solla, and D. A. Cohn, Eds., Cambridge, MA, 1999, pp. 330–336, MIT Press.

[18] J. Martín-Guerrero, G. Camps-Valls, E. Soria-Olivas, A. J. Serrano-López, J. Pérez-Ruixo, and N. V. Jiménez-Torres, “Dosage individualization of erythropoietin using a profile–dependent support vector regression,” *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 10, pp. 1136–1142, October 2003.

[19] G. Camps-Valls, E. Soria-Olivas, J. Pérez-Ruixo, A. Artés-Rodríguez, F. Pérez-Cruz, and A. Figueiras-Vidal, “Cyclosporine concentration prediction using clustering and Support Vector Regression methods,” *IEE Electronics Letters*, vol. 38, no. 12, pp. 568–570, June 2002.

[20] A.B. Watson and J.A. Solomon, “A model of visual contrast gain control and pattern masking,” *Journal of the Optical Society of America A*, vol. 14, no. 9, pp. 2379–2391, September 1997.

[21] P.C. Teo and D.J. Heeger, “Perceptual image distortion,” *Proceedings of the First IEEE International Conference on Image Processing*, vol. 2, pp. 982–986, 1994.

[22] A.B. Watson, “DCT quantization matrices visually optimized for individual images,” in *Human Vision, Visual Processing and Digital Display IV*, Bernice E. Rogowitz; Eds. Jan P. Allebach, Ed. 1993, vol. 1913 of *Proceedings of the SPIE*, pp. 202–216, SPIE–The International Society for Optical Engineering.

[23] I. Epifanio, J. Gutiérrez, and J. Malo, “Linear transform for simultaneous diagonalization of covariance and perceptual metric matrix in image coding,” *Pattern Recognition*, vol. 36, no. 8, pp. 1799–1811, August 2003.

[24] F.W. Campbell and J.G. Robson, “Application of Fourier analysis to the visibility of gratings,” *Journal of Physiology*, vol. 197, no. 3, pp. 551–566, August 1968.

[25] B. Girod, “What’s wrong with mean-squared error,” in *Digital Images and Human Vision*, A. B. Watson, Ed., pp. 207–220. The MIT press, 1993.

[26] J. Malo, A.M. Pons, and J.M. Artigas, “Subjective image fidelity metric based on bit allocation of the human visual system in the DCT domain,” *Image & Vision Computing*, vol. 15, no. 7, pp. 535–548, July 1997.

[27] A. B. Watson and J. Malo, “Video quality measures based on the standard spatial observer,” in *Proceedings of the IEEE International Conference on Image Proceedings*, 2002, vol. 3, pp. 41–44.

[28] K.N. Nygan, K.S. Leong, and H. Singh, “Adaptive cosine transform coding of images in the perceptual domain,” *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 37, no. 11, pp. 1743–1750, November 1989.

LIST OF FIGURES

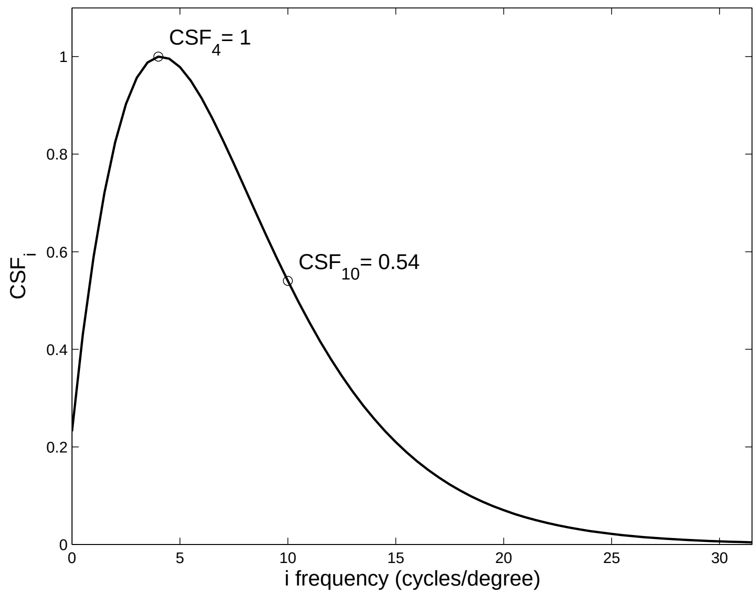


Fig. 1. Contrast Sensitivity Function (CSF) of Nygan et al. [28]. The slopes of two particular sensors respectively tuned to low-frequency stimuli ($\mathrm{CSF}_4 = 1$) and high-frequency stimuli ($\mathrm{CSF}_{10} = 0.54$) have been highlighted.

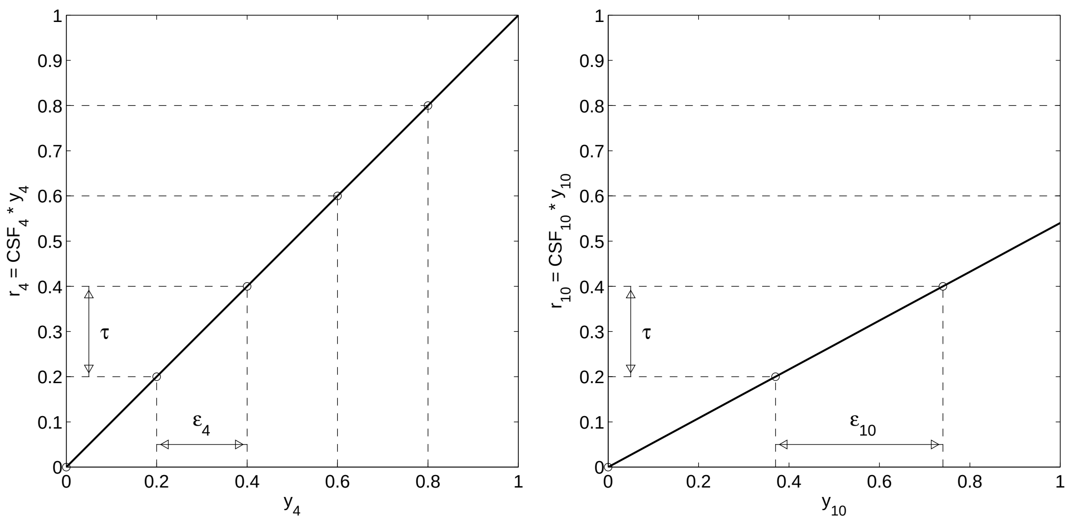


Fig. 2. Responses and associated visibility thresholds (insensitivity regions) of the two sensors whose slopes have been highlighted in Fig. 1. The Euclidean nature of the response domain implies that two distortions, Δy_i and Δy_j, induce perceptually equivalent effects if the corresponding variations in the response are the same: $\Delta r_i = \Delta r_j = \tau$. This is why, assuming a certain threshold for MPE$_s$, the biggest the slope in the response, i, the smallest the acceptable distortion in y_i, giving rise to Eq. (17).

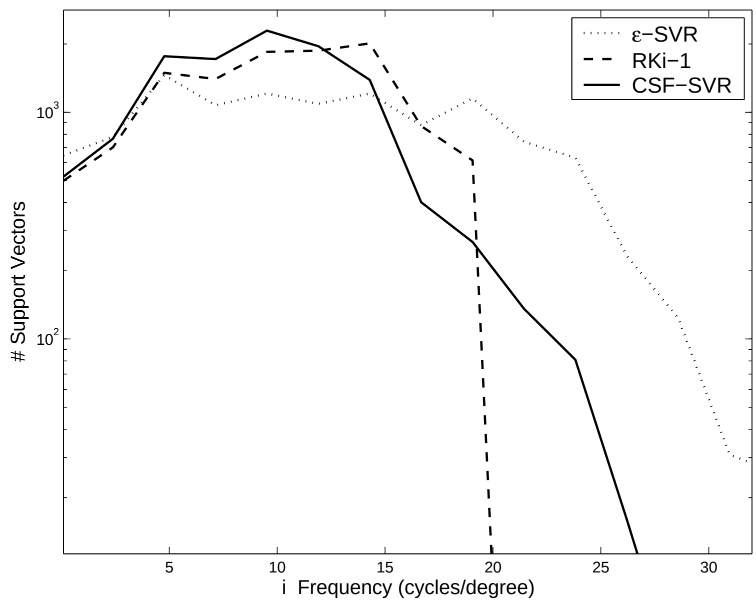
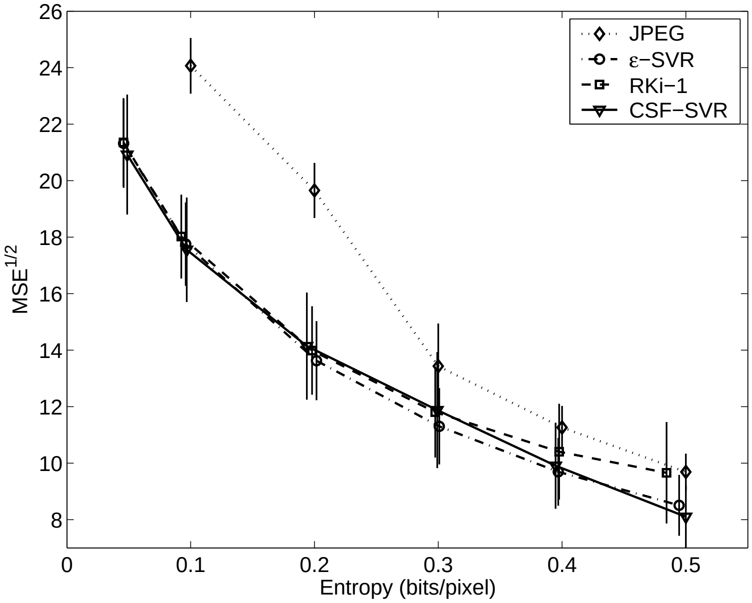


Fig. 3. Distribution of support vectors (SVs) for each ε profile as a function of the frequency in the Lena image.

(a)



(b)

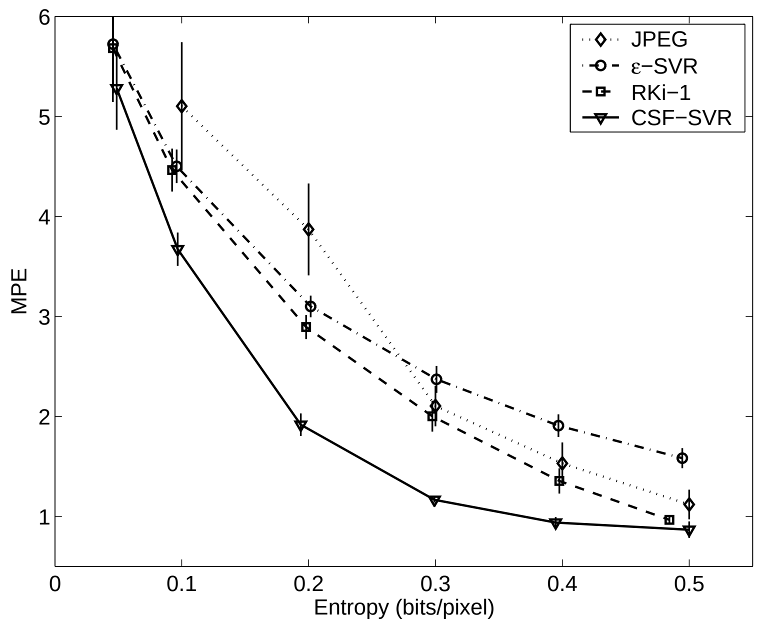


Fig. 4. Rate distortion curves of JPEG and the three SVM-based image coding methods. (a) Distortion measured with the standard MSE$^{1/2}$. (b) Distortion measured using the perceptually meaningful MPE. These results are the average over the five standard images, and the error bars stand for the standard deviation of the corresponding distortion at each point.

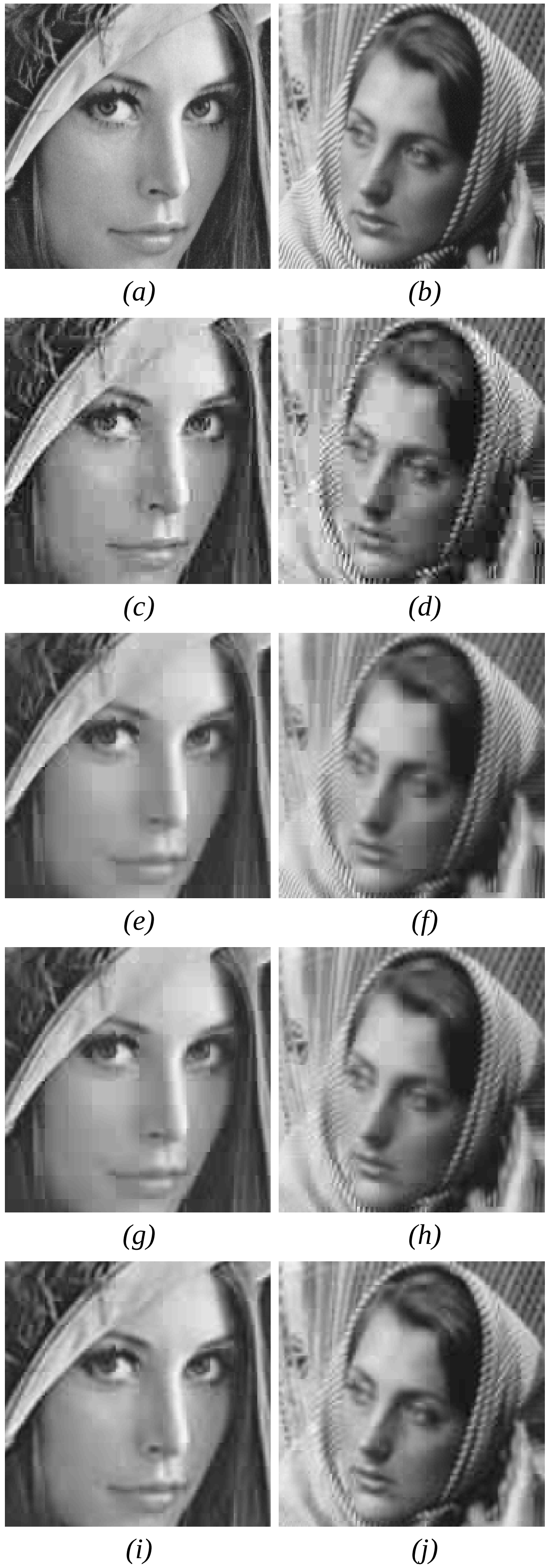


Fig. 5. Examples of decoded images. (a) Lena, and (b) Barbara (zoom of the original images at 8 bits/pixel). The bit-rate for these examples is 0.3 bpp (27:1) (Lena) and 0.4 bpp (20:1) (Barbara). (c) and (d) JPEG, (e) and (f) ε-SVR, (g) and (h) RKi-1, and (i) and (j) CSF-SVR.

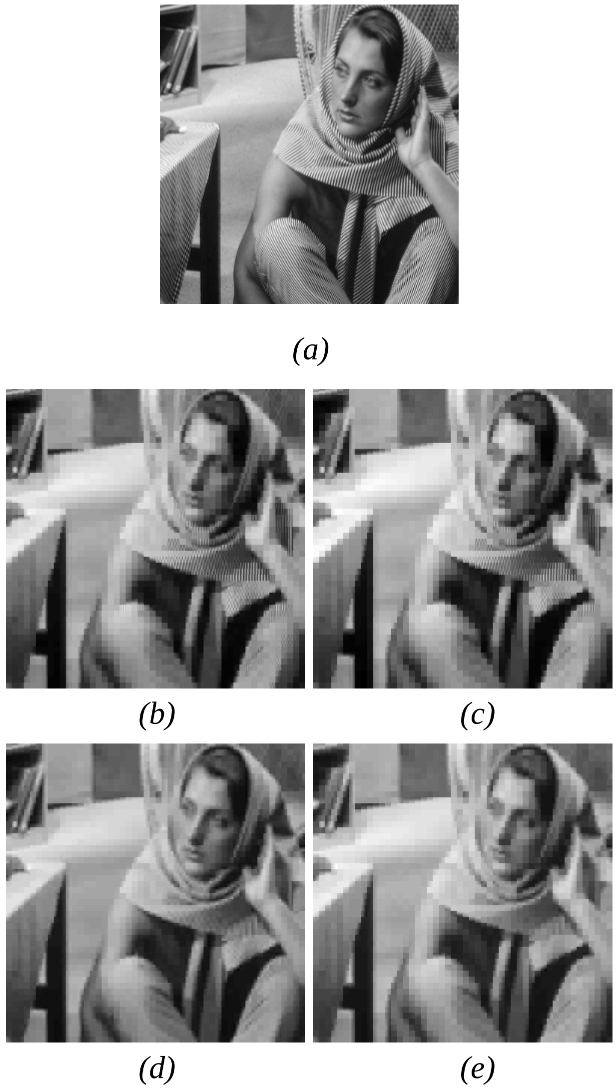
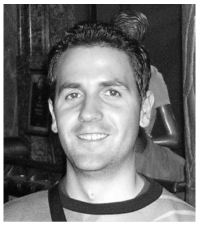


Fig. 6. Examples of decoded images using the RKi-1 and the proposed CSF-SVR training strategies at high compression ratios: 0.1 bpp (64:1) [left] and 0.065 bpp (125:1) [right]. (a) Original Barbara image, (b) and (c) RKi-1, and (d) and (e) CSF-SVR.

Compression ratio	$MSE^{1/2}$		MPE	
	RKi-1	**CSF-SVR**	**RKi-1**	**CSF-SVR**
0.10 bpp (64:1)	17.5	17.4	6.2	5.0
0.08 bpp (100:1)	18.0	17.8	6.6	5.5
0.065 bpp (125:1)	18.7	18.5	7.1	6.4

TABLE I

OBJECTIVE ($MSE^{1/2}$) AND SUBJECTIVE (MPE) ERRORS OF THE DECODED IMAGES AT HIGH COMPRESSION.



Gabriel Gómez (1980) received the B.Sc. degree in Physics in 2002 and the B.Sc. in Electrical Engineering in 2004 both from the Universitat de València.

Currently, he is working at the Grupo de Procesado Digital de Señales (GPDS) and at the Visual Statistics Group (VI(S)TA) of the Universitat de València. He is interested in machine learning algorithms (Support Vector Machines and Reinforcement Learning) and image processing.

Gustavo Camps-Valls (M'04) was born in València, Spain in 1972, and received a B.Sc. degree in Physics (1996), a B.Sc. degree in Electronics Engineering (1998), and a PhD degree in Physics (2002) from the Universitat de València. He is currently an assistant professor in the Department of Electronics Engineering at the Universitat de València, where teaches electronics, advanced time series processing, and digital signal processing. His research interests are neural networks and kernel methods for image processing and biomedical engineering. He is the author (or co-author) of 25 journal papers, several book chapters, and more than 75 international conference papers. He is a referee of several international journals and has served on the Scientific Committees of several international conferences. Visit http://www.uv.es/$\sim$gcamps for more information.

Juan Gutiérrez (1971) received the *Licenciado* degree in Physics (Electricity, Electronics and Computer Science) in 1995 from the Universitat de València.

He is with the Computer Science Department of the Universtitat de València, where he is an Assistant Professor. Currently, he is with the Visual Statistics Group (VI(S)TA) at the Universitat de València (http://www.uv.es/vista/vistavalencia). His current research interests include regularization theory, models of low-level human vision, and representation and analysis of images.

Jesús Malo (1970) received the M.Sc. degree in Physics in 1995 and the Ph.D. degree in Physics in 1999 both from the Universitat de València.

He was the recipient of the Vistakon European Research Award in 1994. In 2000 and 2001 he worked as Fulbright Postdoc at the Vision Group of the NASA Ames Research Center, and at the Lab of Computational Vision of the Center for Neural Science, New York University. Currently, he is with the Visual Statistics Group (VI(S)TA) at the Universitat de València (http://www.uv.es/vista/vistavalencia). He is member of the Asociación de Mujeres Investigadoras y Tecnólogas (AMIT).

He is interested in models of low-level human vision, their relations with information theory, and their applications to image processing and vision science experimentation. His interests also include (but are not limited to) Fourier, Matlab, modern art, independent movies, chamber music, Lou Reed, Belle and Sebastian, The Pixies, comics, Faemino y Cansado, la Bola de Cristal, and beauty in general...