Normalized Image Representation for Efficient Coding

Jesús Malo
Dept. d'Òptica, Facultat de Física, Universitat de València
Dr. Moliner 50. 46100 Burjassot, València, (SPAIN)
jesus.malo@uv.es

Abstract— **In this paper we propose an adaptive non-linear image representation based on the divisive normalization of local-frequency transforms used in contrast masking models. This normalized representation has two effects: (1) it increases the statistical independence of the coefficients of the representation and (2) it is Euclidean from a perceptual point of view.**

Experimental results show that reducing the remaining statistical and perceptual dependence using normalized representations for transform coding may make a big difference in the quality of the reconstructed images.

I. Introduction

The aim of the change of representation in transform coding is twofold [2]: it is intended to remove the *statistical* and the *perceptual* dependence between the image coefficients.

A number of linear transforms (such as PCA, DCT, ICA or Wavelets) have been used to reduce the statistical dependence between the coefficients of the representation [3]–[5]. In the conventional approach to transform coding, perceptual factors are taken into account only *after* the selection of the representation, in the quantizer design. Moreover, in order to apply the standard theory for bit allocation, the (perceptual) metric has to be diagonal in the representation to be quantized [4].

However, the above linear transforms do not completely achieve the desired independence from both points of view [2], [6]. This means that scalar quantization of these representations is not completely appropriate.

It has been shown that using non-linearities in which each coefficient is normalized by a combination of the neighboring coefficients (the local variance [7] or a linear combination of the energy of the neighbors [6]) gives rise to signals with interesting marginal probability density functions and increased independence. This *divisive normalization* non-linearity is the transform that takes place after the local-frequency analysis in biological early vision [8], [9]. Besides, this kind of perceptually inspired normalization naturally leads to a perceptually Euclidean domain [2], [10].

According to the referred (statistical and perceptual) properties of the divisive normalization, it could make a difference in image coding. However, using the divisive normalization is not straightforward since *it is not easily invertible*. This is a critical issue because the reconstructed image has to be obtained inverting the non-linearity from the quantized coefficients.

In this work we propose the use of a psychophysically inspired divisive normalization to obtain an image representation which is perceptually Euclidean and has an increased independence at the same time. We present a computationally efficient method to invert the representation and we analyze the invertibility condition and its robustness to quantization. Finally, we show that removing the remaining dependence in linear transforms using this normalization prior to quantization makes a difference in the quality of the reconstructed images.

II. Divisive normalization models

The current models of early visual processing in the human cortex involve two stages:

$$A \xrightarrow{T} a \xrightarrow{R} r \tag{1}$$

where the image, A, is first analyzed by a (linear) wavelet-like filter bank, T [9], and R is a non-linear transform of the wavelet coefficients: the *divisive normalization* [8], [9]. The linear filter bank, T, leads to a local-frequency representation similar to the one used in transform coding (such as block-DCT or Wavelets). The divisive normalization models describe the gain control mechanisms normalizing the energy of each linear coefficient by a linear combination of its neighbors in space, orientation and scale:

$$r_i = \frac{|a_i|^\gamma}{\beta_i + (h \cdot |a|^\gamma)_i} \tag{2}$$

where h is a matrix that defines the neighborhoods that describe the *masking interactions* between all the coefficients of the vector a, and the rectification (the absolute value) and the exponent γ are applied to each coefficient of the vector a. The sign (or phase) of each coefficient is inherited from the sign of the corresponding linear coefficient.

Note that this scheme is similar to the one used in transform coding, where first a linear transform is used to reduce the statistical dependence between the samples of A and then some additional non-linearity may be considered in order to simplify the quantizer design [3], [4].

In the visual psychophysics context [9], the parameters of the divisive normalization are chosen to fit the experimental contrast incremental thresholds (i.e. the inverse of the slope

This work was partially supported by the projects UV-AE-03-29 and TIC2003-01504 (co-funded by the Spanish CICYT and the European FEDER).

An extended version of this work was submitted to IEEE Trans. Im. Proc. [1].

0-7803-8104-1/03/$17.00 ©2003 IEEE

of the response). As illustration we propose normalization parameters for a particular set of local-frequency basis functions: the block-DCT.

We use the parameters that fit the contrast incremental thresholds of sinusoidal grids measured at our lab. The experimental procedure was similar to the one used in [9], [11]. In this fit we have only considered Gaussian neighborhoods in scale (frequency) and orientation because these particular experiments didn't explore the spatial interactions. This is not a problem in the case of applications that use extended basis functions in each region such as the block-DCT. As in [9], here, an additional scalar weighting parameter, α, is included in the transform, T, simulating the global band-pass response of the filter bank (the *Contrast Sensitivity Function*, CSF). This means that the transform coefficients, a_i, are given by:

$$a_i = \alpha_i \cdot \sum_{j=1}^{N^2} T_{ij}\, A_j$$

Figure 1 shows the values of the parameters, α, β and h that reproduce the contrast incremental threshold data and some examples of the response curves for certain local-DCT patterns. The value of the excitation and inhibition exponent has been fixed to $\gamma = 2$ [8].

For other basis of interest such as wavelets, the normalization model can be extended introducing spatial interactions in the Gaussian kernel, h, using the results reported in [9] or [12]: the spatial extent of the interactions is about twice the size of the impulse response of the CSF.

Given an image, A, of size $N \times N$, if a non-redundant basis is used to model T, the size of the vectors a, r and β is N^2, and the size of h is $N^2 \times N^2$. Considering these sizes, an arbitrary interaction pattern in h would imply an explicit (expensive) computation of the product $h \cdot |a|^\gamma$. Fortunately, the nature of the interactions between the coefficients is *local* [9], [12], as shown in figure 1. This fact induces a sparse structure in h and allows a very efficient computation of $h \cdot |a|^\gamma$ using simple convolutions.

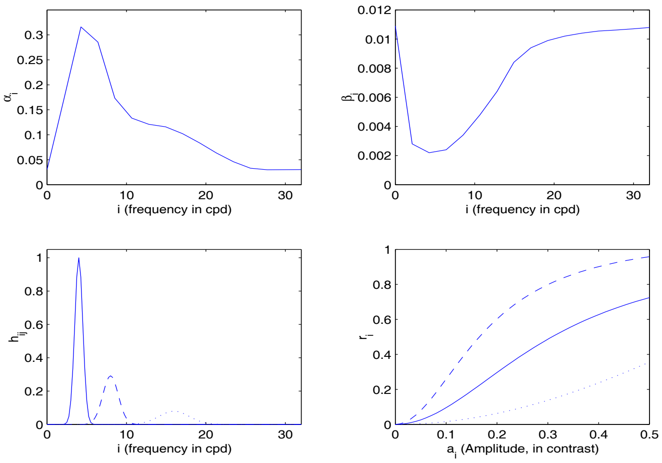

Fig. 1. *Psychophysical parameters α, β and (examples of the interaction kernels in) h for the local-DCT case. The different line styles represent different frequencies: 4 cpd (solid), 8 cpd (dashed) and 16 cpd (dotted). The bottom right figure shows some examples of the response for the corresponding basis functions on a zero background.*

III. Benefits of divisive normalization for image coding

As stated in the introduction the aim of the image representation in the context of transform coding should be reducing the statistical and the perceptual dependence between the coefficients at the same time.

The statistical dependence is usually described by a non-diagonal covariance matrix, Γ [3], [4]. The perceptual dependence may be described by a non-Euclidean perceptual metric, W [2]. The efficiency of a representation from both points of view may be evaluated analyzing the non-diagonal nature of these matrices. This can be measured using a parameter, η_s (for Γ) or η_p (for W), that is defined as the ratio between the magnitude of the off-diagonal coefficients of the (statistical or perceptual) matrix with the magnitude of their diagonal coefficients [3].

Table I shows these interaction measures for the spatial domain, for two classical linear local-frequency domains and for the proposed domain: *local-DCT plus divisive normalization*.

On one hand we have computed the statistical interaction measure, η_s, on the covariance of the samples in the usual way, i.e. *taking into account their sign*. In this case, as expected, the local-PCA, which is designed to diagonalize Γ, achieves the best η_s result. The local-DCT which is a good fixed-basis approximation of the PCA [3], achieves a quite good result as well. However, notice that if the statistical relation between the absolute value of the coefficients is analyzed (measure $\eta_{|s|}$) it is obvious that the linear transforms do not remove these interactions. On the other hand, we see that the proposed image representation does reduce the statistical interactions. And this is true even in the absolute value (or energy) case, which is something that linear transforms cannot do.

Beyond these statistical facts, the non-linear interactions after the transform domain imply that the metric, W, estimated using Riemannian geometry is not diagonal in any linear representation. In particular, the coefficients of the metric in the linear representation given by the filters T in eq. 1 depend on the slope (jacobian) of the non-linear response in eq. 2. As this slope is non-diagonal and input-dependent, the metric, W, cannot be diagonalized in any linear domain (see [1], [2] for details).

The simultaneous consideration of both aspects makes the proposed representation a good candidate for transform coding. The only technical issue to be analyzed before using the

	Spatial Domain	*local-DCT*	*local-PCA*	*Normaliz. Domain*
η_s	169.2	6.6	0.0	0.7
$\eta_{\lvert s \rvert}$	169.2	21.5	17.1	1.3
η_p	48.2	1.1	12.6	0.0

TABLE I

Stat. and Perceptual interaction measures in different domains.

normalization is its inversion in order to come back to the spatial domain after the quantization.

IV. Inversion of the normalization

Let D_r and D_β be diagonal matrices with the vectors r and β in the diagonal, then from eq. 2 it follows:

$$|a|^\gamma = (I - D_r \cdot h)^{-1} \cdot D_\beta \cdot r \tag{3}$$

However, this analytic solution is not practical because of three reasons. First, the matrices are $huge$ so computing the inverse $(I - D_r \cdot h)^{-1}$ is very expensive. Second, while in the normalization the interactions between the coefficients of a are *local* (h is sparse), in the inverse the interactions between the coefficients of r are *global*, i.e., $(I - D_r \cdot h)^{-1}$ is dense. This dense interaction makes eq. 3 hard to use even with moderate size images. And third (the worse one), the inverse will not exist if $I - D_r \cdot h$ is singular.

A. Series expansion inversion

The particular form of the normalization model and the corresponding inverse allows us to propose an alternative solution that doesn't involve matrix inversions nor dense matrices. The idea is using a series expansion of the inverse matrix in eq. 3.

$$(I - D_r \cdot h)^{-1} = \sum_{k=0}^{\infty} (D_r \cdot h)^k$$

In that way, we can compute the inverse up to a certain degree of approximation, taking $n < \inf$ terms in the series:

$$\begin{aligned} |a|^\gamma{}_{(1)} &= D_\beta \cdot r + (D_r \cdot h) \cdot D_\beta \cdot r \\ |a|^\gamma{}_{(2)} &= D_\beta \cdot r + (D_r \cdot h) \cdot D_\beta \cdot r + (D_r \cdot h)^2 \cdot D_\beta \cdot r \\ &\vdots \end{aligned}$$

A naive implementation would imply computing powers of $D_r \cdot h$ which is also a problem. However, taking into account that the previous equations can be rewritten as:

$$\begin{aligned} |a|^\gamma{}_{(1)} &= D_\beta \cdot r + (D_r \cdot h) \cdot D_\beta \cdot r \\ |a|^\gamma{}_{(2)} &= D_\beta \cdot r + (D_r \cdot h) \cdot ((D_r \cdot h) \cdot D_\beta \cdot r + D_\beta \cdot r) \\ &\vdots \end{aligned}$$

we can write the series approximation in a recursive fashion that only involves vector additions and matrix-on-vector multiplications:

$$\begin{aligned} |a|^\gamma{}_{(0)} &= D_\beta \cdot r \\ |a|^\gamma{}_{(n)} &= D_\beta \cdot r + D_r \cdot h \cdot |a|^\gamma{}_{(n-1)} \end{aligned} \tag{4}$$

Note that the matrices in eq. 4 are sparse and therefore it allows a fast implementation using convolutions.

B. Invertibility and convergence condition

The same condition has to hold to ensure the existence of the solution and the convergence of the series inversion method. Let V and λ be the eigenvector and eigenvalue matrix decomposition of $D_r \cdot h$:

$$D_r \cdot h = V \cdot \lambda \cdot V^{-1}$$

As we show below, the invertibility condition turns out to be:

$$\lambda_{max} = max(\lambda_i) < 1 \tag{5}$$

In the *analytic* case the matrix $(I - D_r \cdot h)$ has to be invertible, i.e. $det(I - D_r \cdot h) \neq 0$. However if some eigenvalue, λ_i, is equal to one, then $det(\lambda_i I - D_r \cdot h) = 0$. In theory, it would be enough to ensure that $\lambda_i \neq 1$, but in practice, as the spectrum of $D_r \cdot h$ is almost continuous (see the example in figure 2), it is very likely to have *dangerous* eigenvalues if the condition 5 doesn't hold.

In the *series expansion* case, the convergence of the series has to be guaranteed. Using the eigenvalue decomposition of $D_r \cdot h$ in the expansion, we find:

$$\sum_{k=0}^{\infty} (D_r \cdot h)^k = V \cdot \left(\sum_{k=0}^{\infty} \lambda^k \right) \cdot V^{-1}$$

which clearly converges only if the maximum eigenvalue is smaller than one.

We have empirically checked the invertibility of the *psychophysically inspired* normalization for the local-DCT case by computing the maximum eigenvalue of $D_r \cdot h$ over the blocks of a set of 200 images of the Van Hateren natural image data set [13]. Figure 2 shows the average eigenvalues spectrum with the corresponding standard deviation. As the obtained eigenvalues are smaller than 1, the normalization with these psychophysical parameters will be invertible. Besides, as they are *far enough* from 1 it will remain invertible even if the responses undergo small distortions such as quantization.

C. Convergence rate

It is possible to derive an analytic description for the convergence of the *series expansion* method. It turns out that

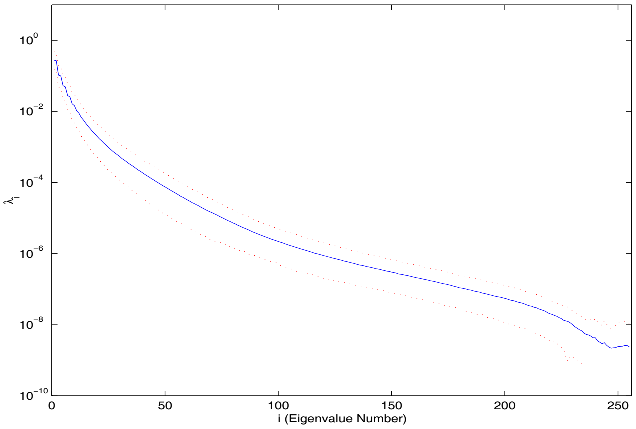

Fig. 2. *Average eigenvalues of $D_r \cdot h$ for 200 natural images.*

the convergence is faster for a smaller λ_{max}. Consider that the error vector at the step n of the approximation,

$$e_{(n)} = |a|^{\gamma} - |a|^{\gamma}_{(n)}$$

is just the last part of the series, and using the eigenvalue decomposition of $D_r \cdot h$, we have:

$$e_{(n)} = \sum_{k=n+1}^{\infty} (D_r \cdot h)^k \cdot D_\beta \cdot r = \sum_{k=0}^{\infty} (D_r \cdot h)^{(n+k+1)} \cdot D_\beta \cdot r$$
$$= V \cdot \left(\sum_{k=0}^{\infty} \lambda^{(n+k+1)} \right) \cdot V^{-1} \cdot D_\beta \cdot r$$

Then, taking the $|\cdot|_\infty$ norm as a measure of the error, we have:

$$\begin{aligned} \epsilon_{(n)} = |e_{(n)}|_\infty &= max(e_{(n)\,i}) \propto \sum_{k=0}^{\infty} {\lambda_{max}}^{(n+k+1)} \\ &= \sum_{m=0}^{\infty} ({\lambda_{max}}^n)^m - 1 \end{aligned}$$

and therefore, the error at each step is:

$$\epsilon_{(n)} \propto \frac{{\lambda_{max}}^n}{1 - {\lambda_{max}}^n} \qquad (6)$$

Figure 3 confirms this convergence rule: it shows the evolution of the error measure as a function of the number of terms in the series for three images with different λ_{max}. From eq. 6 it follows that for a big enough number of terms it holds $\log(\epsilon_{(n)}) \propto \log(\lambda_{max}) \cdot n$, as shown in the figure. The experiment in figure 3 shows the result of local-DCT blocks, but the same behavior is obtained in the wavelet case.

D. Robustness to quantization

Figure 4 shows the effect of the quantization step (number of bits per coefficient) on λ_{max} which is the key for the invertibility (and convergence). These results capture the evolution of the maximum eigenvalue of data set used in figure 2 when compressing the images in the range 1–0.02 bits/pix. For relatively high bit-rates (over 0.6 bits/pix) the maximum eigenvalue remains stable and equal to its value in the original signal. For smaller bit-rates λ_{max} oscillates a little bit, but it always lays in the region that allows the invertibility.

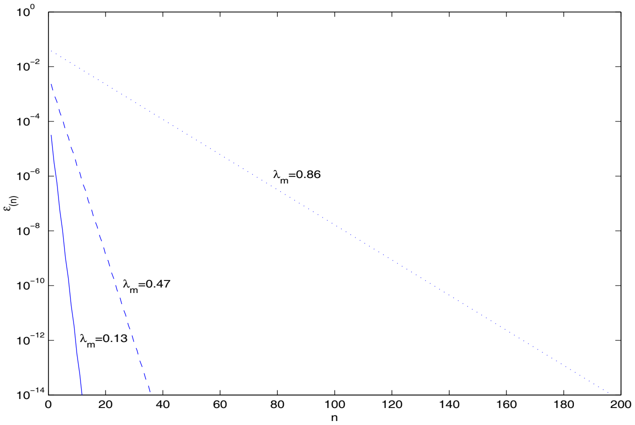

Fig. 3. *Error the series expansion method as a function of the number of terms in the series. The different lines represent the error obtained when inverting different images with different values of λ_{max}.*

This ensures that the proposed normalized representation is invertible no matter the bit-rate: the coarseness of the quantization is not limited by the invertibility condition but just by the admissible distortion (as usual).

V. Image coding results

The nature of the quantization noise depends on the quantizer design. The quantizers based on the minimization of the MSE end with non-uniform quantization solutions based on the marginal PDFs [4] or some modification of them including the perceptual metric [2], [14]. However, it has been suggested that constraining the Maximum Perceptual Error (MPE) may be better than minimizing its average [14]. This is because the important issue is not minimizing the average error across the regions, but minimizing the annoyance in every region.

Constraining the MPE is equivalent to a uniform quantization in a perceptually uniform domain. Therefore, *once in the perceptually Euclidean domain* the quantizer design is extremely simple: uniform scalar quantizers and uniform bit allocation. Of course, the expression of this quantizer turns out to be non-uniform in the (intermediate) transform domain (local-DCT or wavelets).

The difference between the approaches that implicitly followed the MPE idea [2], [14]–[18] is the accuracy of the perception model which is used to propose the perceptually Euclidean domain. For instance, the quantization scheme (empirically) recommended in the JPEG standard [16] may be deduced from the MPE restriction with a very simple linear vision model based on the CSF [14]. In this case, the perceptual metric is fixed (the model is linear) and it is assumed to be diagonal in the local DCT domain. In this very simple case, it is assumed that no perceptual relationship exists between the

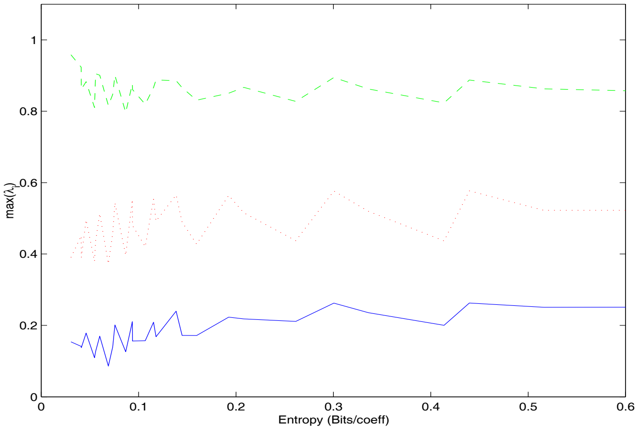

Fig. 4. *Effect of quantization on λ_{max}. The solid line and the dotted line stand for the average and the average plus one standard deviation of λ_{max} over the considered data set. The dashed line shows the behavior of the maximum λ_{max} of the data set.*

coefficients of the transform, and that the perceptual relevance of each coefficient is given by the corresponding CSF value. The performance of this approach can be improved at around 0.5 bits/pix if a more sophisticated model is used [14], [15], [17], [18]. In these references the authors used a point-wise non-linear model in the DCT domain. In that case the metric is image adaptive but still it is assumed to be diagonal, i.e., no interactions are considered between coefficients. These results can be improved if the perceptual interactions are not neglected [2]. In that case, the authors used a fixed (average) non-diagonal metric together with the correlation matrix in order to represent the signal in a perceptually and statistically decorrelated domain. However, in this case the authors had to neglect the input adaptive behavior because of the inversion problem. In this paper we follow the same MPE (uniform quantization approach) but using the state-of-the-art perceptual model: the divisive normalization model that implies a non-diagonal and input-dependent metric.

Here (figure 5 and table II) we compare the results of different MPE transform coding schemes at the same compression ratio using image representations (or vision models) of progressively increasing accuracy: JPEG [16] that assumes the linear CSF model and a fixed diagonal metric in the DCT domain (fig. 5, top-left), the algorithm of Malo et al. [14], [18], that assumes a point-wise non-linear model [11] and hence an input-dependent diagonal metric in the DCT domain (fig. 5, top-right), the algorithm of Epifanio et al. [2], that assumes a fixed non-diagonal metric in the DCT domain (fig. 5, bottom-left), and the proposed representation, that uses the psychophysical normalization assumes an input-dependent and non-diagonal perceptual metric (fig. 5, bottom-right).

VI. Conclusion

Image coding results suggest that a straightforward uniform quantization of the normalized coefficients is a promising alternative to the current transform coding techniques that use different degrees of perceptual information in the image representation and quantizer design. These results show that removing or reducing the remaining (statistical and perceptual) dependence in linear transforms may make a big difference in the quality of the reconstructed images.

	Fixed diag. W	Adaptive diag. W	Fixed non-diag. W	Adaptive non-diag.W
MSE (0.28)	257.8	229.8	177.1	100.2
PMSE	100.2	129.5	44.2	30.5
MSE (0.43)	240.5	197.8	148.4	67.7
PMSE	95.8	90.1	33.4	23.4
MSE (0.58)	224.8	156.6	118.7	47.1
PMSE	70.7	63.0	19.6	5.3

TABLE II

Objective (MSE) and subjective (Perceptual MSE [2], [14], [18]) distortions at different bit rates (in parenthesis, in bit/pix).

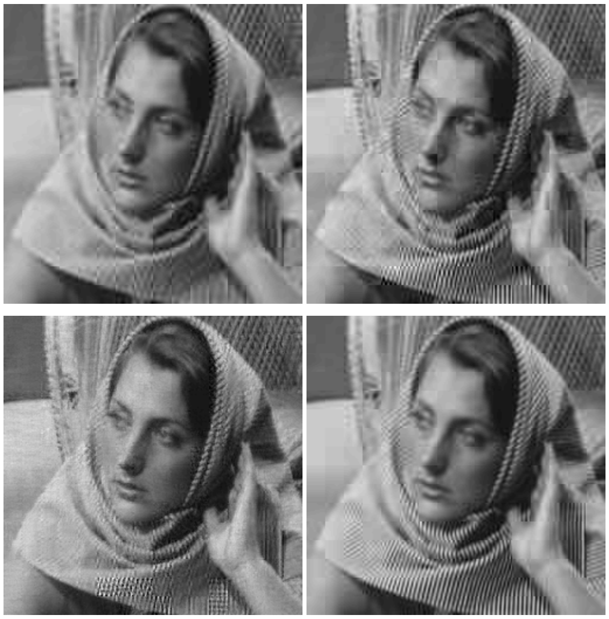

Fig. 5. *Coding results on the Barbara image at 0.28 bits/pix. JPEG (top-left), MPE quantizer using the CSF (fixed diagonal W) [16]. MPE quantizer using a point-wise non-linearity (adaptive diagonal W) [14], [18] (top-right). MPE quantizer using a fixed non-diagonal W [2] (bottom-left). MPE quantizer in a normalized domain, i.e., adaptive non-diagonal W (bottom-right).*

References

[1] J. Malo, E. Simoncelli, I. Epifanio, and R. Navarro. Non-linear image representation for efficient coding. *Submitted to IEEE Trans. Im. Proc.*, 2003.

[2] I. Epifanio, J. Gutiérrez, and J.Malo. Linear transform for simultaneous diagonalization of covariance and perceptual metric matrix in image coding. *Pattern Recognition*, 36:1799–1811, 2003.

[3] R.J. Clarke. *Transform Coding of Images*. Acad. Press, New York, 1985.

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

[5] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, New York, 2001.

[6] R.W. Buccigrossi and E.P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Transactions on Image Processing*, 8(12):1688–1701, 1999.

[7] D.L. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73(6):814–817, 1994.

[8] D. J. Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198, 1992.

[9] A.B. Watson and J.A. Solomon. A model of visual contrast gain control and pattern masking. *JOSA A*, 14:2379–2391, 1997.

[10] P.C. Teo and D.J. Heeger. Perceptual image distortion. *Proceedings of the SPIE*, 2179:127–141, 1994.

[11] G.E Legge. A power law for contrast discrimination. *Vision Research*, 18:68–91, 1981.

[12] A.B. Watson and J.Malo. Video quality measures based on the standard spatial observer. *Proc. IEEE Intl. Conf. Im. Proc.*, 3:41–44, 2002.

[13] J.H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc.R.Soc.Lond. B*, 265:359–366, 1998.

[14] 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*, 18(3):233–246, 2000.

[15] S. Daly. Application of a noise-adaptive Contrast Sensitivity Function to image data compression. *Optical Engineering*, 29(8):977–987, 1990.

[16] G.K. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, 34(4):31–43, 1991.

[17] A.B. Watson. DCT quantization matrices visually optimized for individual images. In B.E. Rogowitz, editor, *Human Vision, Visual Processing and Digital Display IV*, volume 1913, 1993.

[18] J.Malo, J.Gutierrez, I.Epifanio, F.Ferri, and J.M.Artigas. Perceptual feed-back in multigrid motion estimation using an improved DCT quantization. *IEEE Trans. Im. Proc.*, 10(10):1411–1427, 2001.