# Non-Linear Image Representation for Efficient Perceptual Coding

J. Malo, I. Epifanio, R. Navarro, E.P. Simoncelli

***Abstract*— Image compression systems commonly operate by transforming the input signal into a new representation whose elements are independently quantized. The success of such a system depends on two properties of the representation. First, the coding rate is minimized only if the elements of the representation are *statistically* independent. Second, the perceived coding distortion is minimized only if the errors in a reconstructed image arising from quantization of the different elements of the representation are *perceptually* independent. We argue that linear transforms cannot achieve either of these goals, and propose instead an adaptive non-linear image representation in which each coefficient of a linear transform is divided by a weighted sum of coefficient amplitudes in a generalized neighborhood. We then show that the divisive operation greatly reduces both the statistical and the perceptual redundancy amongst representation elements. We develop an efficient method of inverting this transformation, and we demonstrate through simulations that the dual reduction in dependency can greatly improve the visual quality of compressed images.**

***Index Terms*— Transform Coding. JPEG. Independent Components. Statistical Independence. Perceptual Independence. Scalar Quantization. Non-linear Response. Perceptual Metric.**

## I. Introduction

EFFICIENT encoding of signals relies on an understanding of two fundamental quantities, commonly known as *rate* and *distortion*. The rate expresses the cost of the encoding (typically in bits) and the distortion expresses how closely the decoded signal approximates the original. A large body of literature has shown that the problem can be made much more tractable by transforming the image from an array of pixels into a new representation in which rate or distortion are more easily quantified and controlled. Such transforms are typically linear and in recent years are almost always based on basis functions that provide a local representation of orientation and spatial frequency.

In this paper, we examine a non-linear transformation, motivated by both the statistical properties of typical photographic images and the known properties of the human visual system. The transformation is formed as a cascade of a linear transform and a divisive normalization procedure, in which each transform coefficient is divided by a signal computed from the magnitudes of coefficients of similar spatial position, orientation and frequency. We argue that this representation provides an effective representation for simultaneous optimization of both rate and perceptual distortion in compression of visual images. We begin by reviewing the literature about image statistics and perception leading to the idea of divisive normalization (section II). Section III provides a technical definition of the normalization, explains a particular way to obtain its parameters and illustrates its statistical and perceptual benefits for image coding[1]. In section IV we discuss in detail the problem of inverting a divisive normalization transformation: first we describe a numerical alternative to the analytical inversion, then we derive the general condition for the normalization to be invertible, and finally we check the invertibility of the particular proposed normalization according to this general condition when the coefficients are quantized. In section V we show through simulations that the quality of images reconstructed from the quantized normalization representation can significantly exceed that of images reconstructed from the quantized linear representation.

## II. Background: statistical and perceptual dependence

Traditional transform coding solutions emphasize rate optimization, by attempting to represent the image in a domain where the statistical dependence between coefficients is eliminated. Under this condition, each coefficient may be encoded independently. More specifically, statistical independence of the coefficients justifies the use of scalar quantization and zero-order entropy coding [2, 3]. The classical solution to the problem of transform design is derived by considering only the second-order statistics of the input signal. In this case, the linear transformation that minimizes the correlation of the coefficients may be computed using Principal Components Analysis (PCA). If one assumes spatial stationarity, the Fourier transform achieves this optimum. But this solution is not unique and only considers second-order relations. In recent years a variety of approaches, known collectively as "Independent Components Analysis" (ICA), have been developed to exploit higher-order statistics for the purpose of achieving a

During the development of the work, JM was partially supported by the projects TIC2003-01504 (CICYT-FEDER) and Grupos04-08 (GV), and by the MEC-Fulbright Fellowship FU2000-0029167406. IE was supported by the projects BSA2001-0803-C02-02 (CICYT-FEDER) and GV04B/32 (GV). JM and EPS were partially supported by the Howard Hughes Medical Institute.

JM is with the Department d'Òptica, Universitat de València, 46100 Burjassot, València, Spain (jesus.malo@uv.es, http://www.uv.es/vista/vistavalencia/). IE (epifanio@mat.uji.es) is with Departament de Matemàtiques, Universitat Jaume I, Castelló, Spain. RN is with Instituto de Óptica, CSIC, Spain. EPS is with Howard Hughes Medical Institute, Center for Neural Science and Courant Institute of Mathematical Sciences, New York University, USA. (eero.simoncelli@nyu.edu, http://www.cns.nyu.edu/∼lcv/)

[1]It is important to stress that the statistical benefits of using this advanced perceptual model are not limited to its application in coding: see [1] for an application in image restoration.

unique linear solution [4]. The basis functions obtained when these methods are applied to images are spatially localized, and selective for orientation and spatial frequency (scale) [5, 6], and are thus similar to basis functions of multi-scale wavelet representations.

Despite its name, ICA does *not* actually produce statistically independent coefficients when applied to photographic images. Intuitively, independence would seem unlikely, since images are not formed from linear superpositions of independent patterns: the typical combination rule for the elements of an image is *occlusion*. This suggests that achieving independence requires the introduction of non-linearities in the transform. Empirically, the coefficients of orthonormal wavelet decompositions of visual images are found to be fairly well decorrelated (i.e., their covariance is almost zero). But the amplitudes of coefficients at nearby spatial positions, orientations, and scales are highly correlated [7]. These relationships have been exploited, both implicitly [8, 9] and explicitly [10] in compression systems.

The dependencies between responses of linear filters may be substantially reduced by a non-linear operation known as *divisive normalization*, in which each coefficient is divided by a Minkowski combination of neighboring coefficient amplitudes [7, 10, 11]. This empirical observation is consistent with a hidden Markov model in which the amplitude of groups of coefficients is modulated by a hidden scaling variable [12–16].

The second fundamental ingredient of the transform coding problem is distortion. When coding visual images, distortion should be measured perceptually. Ideally, we would like to be able to express the overall perceived image distortion as an additive combination of the distortions arising from each of the transformed elements, as assumed in the standard theory [2, 3]. This requirement implies that the transformed elements should be perceptually independent: the visibility of the distortion in the image introduced by altering one element should not depend on the values of the other elements. Thus, we should seek a transformation that eliminates perceptual redundancies [11, 17].

The most standard measure of distortion is mean squared error (MSE), computed by averaging the squared intensity differences of distorted and reference image pixels, along with the related quantity of peak signal-to-noise ratio (PSNR). These are appealing because they are simple to calculate, have clear physical meanings, and are mathematically convenient in the context of optimization. But it is well-known that they do not provide a good description of perceived distortion [18–21]. In particular, the visibility of error in one pixel clearly depends on the values of surrounding pixels.

A simple and widely used improvement comes from incorporating the known sensitivity of human vision to different spatial frequencies. Specifically, within the Fourier domain, distortion is measured by summing the squared errors in each frequency, weighting each term by the sensitivity to its corresponding frequency. The most widely known image and video coding standards (JPEG and MPEG) use a block-DCT decomposition to decorrelate the coefficients, and a frequency-dependent quantizer based on the human Contrast Sensitivity Function (CSF) [22, 23]. Similar methods are applied to wavelet image representations such as in JPEG2000 [24]. Note that in all these situations, the perceptual factors are taken into account only after the selection of the representation (e.g., in the quantizer).

It is well known that the perception of errors in coefficients of local frequency or wavelet representations is not independent, a phenomenon known in the perceptual literature as *masking* [25]. Specifically, the presence of large coefficients can reduce the visibility of errors in coefficients that are nearby in position, orientation and scale. The linear coefficients may be modified so as to more accurately represent perceptual distances by *normalizing* (dividing) each coefficient by a gain signal obtained from a combination of adjacent coefficients [11, 17, 18, 25, 26]. This is consistent with recent models of neurons in visual cortex, in which primarily linear neural responses are modulated by a gain signal computed from a combination of other neural responses [27–29].

One can see from this brief description that there has been a remarkable parallel development of transformations that reduce either statistical or perceptual redundancy, beginning with global frequency-based representations, to local frequency or wavelet-based representations, to most recent solution of divisively normalized representations. Perhaps this is not so surprising given that the human visual system is hypothesized to have been shaped, through processes of evolution and development, by the statistical properties of the visual world (for review, see [30]). Although both the statistical and perceptual observations that lead to normalized representations have been exploited in image coding, they have been used indirectly [10, 17]. The fact that normalized representations appear to be the current best choice for reduction of both statistical and perceptual dependencies suggest that one should explicitly encode the normalized local frequency coefficients. In the following sections, we propose an invertible psychophysically-inspired divisive normalization scheme, whose elements are (pairwise) perceptually independent with low statistical dependence. In order to do this, we need to develop an invertible normalization transformation, and must ensure that this inversion process may be used to obtain the decoded image from a set of quantized coefficients.

## III. The divisive normalization model

We define a general divisive normalization as a cascade of two transformation stages:

$$\{a_i\} \xrightarrow{T} \{c_i\} \xrightarrow{R} \{r_i\}, \tag{1}$$

where the image pixels, $\{a_i\}$, are first analyzed using a linear transform $T$, followed by a non-linear transform, $R$, of the linear coefficients [25–29]. The linear transform should be a local-frequency representation as is commonly used in transform coding (e.g., block-DCT or a wavelet filterbank). The divisive normalization stage describes the gain control mechanisms normalizing the energy of each linear coefficient by a linear combination of its neighbors in space, orientation and scale:

$$r_i = \frac{\mathrm{sgn}(c_i) \quad |c_i|^\gamma}{\beta_i + \sum_j h_{ij} |c_j|^\gamma}. \tag{2}$$

Each coefficient $c_i$ is first rectified and exponentiated. Each of the resulting values are then divided by a weighted sum of the others, where $h_{ij}$ is the set of weights that specify the interactions between all the coefficients of the vector $c$ and coefficient $c_i$. The sign (or phase, in the case of a complex-valued transform) of each normalized coefficient, is inherited from the sign of the corresponding linear coefficient, $\mathrm{sgn}(c_i)$.

### A. Model parameters

For this paper, we use a $16 \times 16$-point block DCT for the transformation $T$, in order to facilitate comparisons with the JPEG standard and related literature [17, 22, 23, 31–35]. The main results are general, and would apply to wavelet-style filterbank representations as well, where they are likely to yield better compression results[2].

There are three basic sources from which one can obtain the normalization parameters: psychophysics [25, 26, 37], electrophysiology [29] and image statistics [10, 38, 39]. In the psychophysically-inspired divisive normalization proposed in this paper, the parameters are chosen by fitting data from human vision experiments, using a method similar to that of [25, 37]. As in [25], we augment the standard DCT with an additional scalar weighting parameter, $\alpha$, accounting for the global sensitivity to the frequency range represented by each basis function (the CSF [40]). Thus, the transform coefficients, $c_i$, are given by:

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

where $T_{ij}$ are the basis functions of the linear transform that analyzes the image $a_j$. The amplitudes of the DCT are expressed as contrast values, by dividing the coefficients by the local luminance. Similar contrast measures have been proposed in the literature in the context of pyramidal decompositions [41, 42].

The parameters of the normalization are determined by fitting the slopes of the normalization function in Eq. (2) to the inverses of the psychophysically measured contrast incremental thresholds for gratings [25, 37]. The values of $\alpha$, $\beta$ and $h$ that fit the experimental responses of isolated sinusoidal gratings [43] are shown in Fig. 1. In the same way, the exponent was found to be $\gamma = 0.98$.

Given an image, $a$, of size $N \times N$, if $T$ corresponds to a non-redundant basis, the size of the vectors $c$, $r$, $\alpha$, $\beta$ is $N^2$. The size of the matrix, $h_{ij}$, is $N^2 \times N^2$. For redundant bases the dimensions will be bigger. Considering these sizes, an arbitrary interaction pattern in the matrix $h$ would imply an explicit (expensive) computation of the products $\sum_j h_{ij}|c_j|^\gamma$. As shown in Fig. 1, the nature of the interactions between the coefficients is *local* [25, 44], which means that $h$ need only describe relationships between coefficients of similar spatial frequency and orientation. This fact induces a sparse structure in $h$ and allows a very efficient computation of $\sum_j h_{ij}|c_j|^\gamma$ as a convolution. Since our experimental data don't constrain the shape of the interaction function, we follow [25] and assume that each row of the matrix $h$ has a two-dimensional circular Gaussian form in the Fourier domain. Specifically, we set the kernels $h_{ij}$ as,

$$h_{ij} = exp(-|f_i - f_j|^2/\sigma_{f_i}^2), \quad (3)$$

$$\sigma_{f_i} = \frac{1}{6}|f_i| + 0.05, \quad (4)$$

where $f_i$ and $f_j$ are two-dimensional frequency vectors (with components in cycles per degree) of the $i$th and $j$th basis functions, respectively.

For other bases of interest such as wavelets, the perceptual normalization model can be extended by introducing spatial interactions in the Gaussian kernels. Previous work indicates that the spatial extent of the interactions should be about twice the size of the impulse response of the CSF [25, 44]. See [36, 39] for examples of this kind of kernels in wavelets and ICA.

### B. Perceptual and statistical independence

In this section we describe the perceptual and statistical dependence problems of linear local frequency representations and demonstrate that normalization reduces these problems. First, consider the perceptual dependence. As stated in section II, the coefficients of a representation are perceptually independent if the visibility of the distortion introduced by altering one coefficient doesn't depend on the values of the other coefficients. A quantitative measure of this can be defined using the *perceptual metric matrix* of the representation [17]. Specifically, we write a second-order approximation of the perceptual difference between an image, $a_0$, and a distorted version, $a_0 + \Delta a$ as:

$$d(a_0, a_0 + \Delta a)^2 = \Delta a^T \cdot W_a(a_0) \cdot \Delta a = \sum_i W_a(a_0)_{ii} \Delta a_i^2 + 2 \sum_{i \neq j} W_a(a_0)_{ij} \Delta a_i \Delta a_j. \quad (5)$$

We refer to $W_a(a_0)$ as the perceptual metric matrix in the spatial domain at the point (image) $a_0$. In general, the diagonal elements of such a perceptual metric matrix represent the independent contribution of the deviations in each element to the perceptual distortion, whereas the off-diagonal elements represent the distortion resulting from perceptual interactions between elements. As such, perceptual independence of a representation is equivalent to diagonality of the perceptual metric of that representation.

The perceptual metric for any representation can be computed from that of another representation by transforming according to the rules of Riemannian geometry [45]. If we assume that the normalized domain is perceptually independent (i.e., the matrix is diagonal), as is common for psychophysically-defined normalization models [25, 37] and as suggested by the success of a number of recent image distortion measures [18, 19, 44, 46, 47], then the metric matrix for any linear representation *cannot* be diagonal. To see this, note that in any linear representation, $c'$, defined by $c' = T'^{-1} \cdot c$, the perceptual metric at the point, $c'_0$, is given by [17],

$$W_{c'}(c'_0) = T'^T \cdot \nabla R(c_0)^T \cdot D \cdot \nabla R(c_0) \cdot T', \quad (6)$$

[2]Preliminary comparisons of the proposed method with JPEG2000 show that this is the case [36].

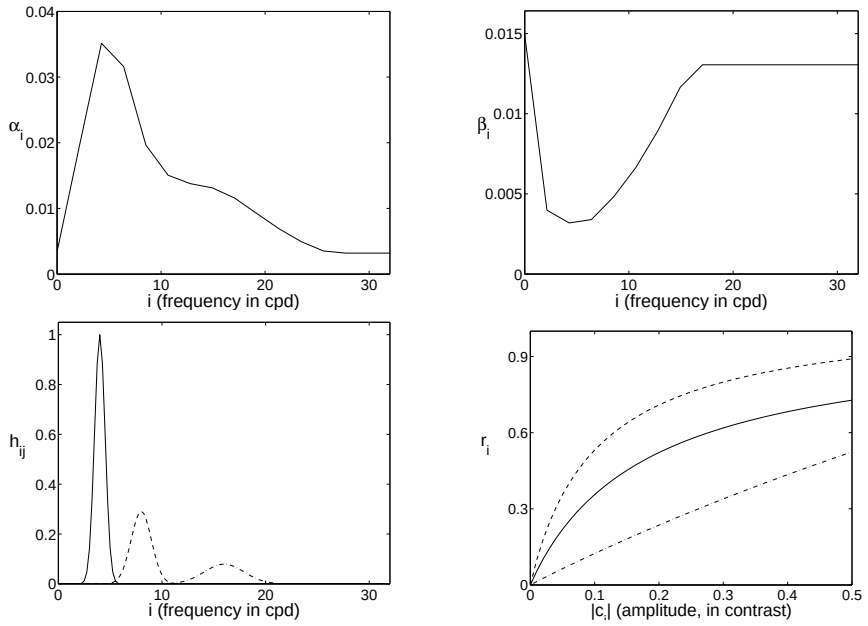



Fig. 1. Parameters $\alpha$, $\beta$ and three interaction kernels (rows of $h$) that fit the contrast incremental threshold data for the DCT case. The different line styles represent different frequencies: 4 cpd (solid), 8 cpd (dashed) and 16 cpd (dash-dot). The bottom right figure shows some examples of the normalized response as a function of coefficient amplitude, on a zero background. Note that the parameters are slightly different from those reported in [1] because we are using here a local DCT instead of a local Fourier transform and a slightly different model. However, the final behavior (bottom right panel) is the same.

where $c_0 = T' \cdot c'_0$, $D$ is the diagonal metric in the normalized domain, and the Jacobian of the transformation is

$$\nabla R(c)_{ij} = \frac{\partial R(c)_i}{\partial c_j} = \\ \operatorname{sgn}(c_i)\gamma \left( \frac{|c_i|^{\gamma-1}\delta_{ij}}{\beta_i + \sum_j h_{ij}|c_j|^\gamma} - \frac{|c_i|^\gamma |c_j|^{\gamma-1} h_{ij}}{(\beta_i + \sum_j h_{ij}|c_j|^\gamma)^2} \right) \quad (7)$$

Assuming the Jacobian is non-diagonal because of the masking interactions ($h_{ij} \neq 0$), and input dependent, no linear representation, $c'$, can achieve the desired perceptual independence.

As described in section II, despite the (second-order) decorrelation power of linear local frequency transforms, their coefficients still exhibit intriguing statistical relationships. A reason for this is that, in general, natural images do not come from a linear combination of signals drawn from independent sources (the central assumption in ICA theory). This means that although the linear representations used in transform coding (the analogue to transform $T$ in the model of Eq. (1)) constitute an appropriate step in the right direction, additional processing is needed to remove (or reduce) the higher-order relations.

As a quantitative measure of the higher-order statistical dependencies, we first use both the cross-correlation and the covariance between the *amplitudes* (absolute values) of the coefficients of the local-DCT representation of a set of images. Second order relationships between the amplitudes (or analogously the energies) correspond to higher-order relationships between the original coefficients. And even in the case of a local frequency transform that is designed to remove the second order correlations in the original signal (e.g. local-PCA and its fixed basis approximation, the local-DCT [2, 48]), the coefficient amplitudes (or energies) may still exhibit strong correlations [7, 10, 11]. Thus, we can use a simple (second order) analysis of the amplitudes of the coefficients as an indicator of independence (or lack of independence) in a broader sense than second-order decorrelation.

First in Fig. 2 we analyze the cross-correlation between the amplitudes of the coefficients of the local DCT transform. As local-DCT spectrum of natural images is not stationary, a direct comparison between coefficients at very different frequencies is biased. Natural images exhibit $\frac{1}{f}$ amplitude spectrum, then, the comparison of a high frequency coefficient with a low frequency coefficient is biased by the high amplitude of the latter. Therefore, instead of a direct comparison, we first divide each coefficient by the average amplitude of that frequency (gathered across all DCT blocks). In that way, a unit mean process is obtained and a fair computation of the cross-correlation can be done. Figure 2 shows cross-correlation contours for amplitudes of nine particular coefficients of increasing frequency in the vertical, horizontal and the diagonal directions. For each of the nine chosen coefficients, the cross-correlation function is maximal at the frequency coordinates of that coefficient, and decreases monotonically as one moves away from those coordinates. This observation is consistent with those reported in other local frequency transform domains [1, 7, 10, 14, 38].

This means that even though local-frequency transforms do remove second order relations, the absolute value of the coefficients is still highly correlated with its neighbors, with an interaction neighborhood that increases in size with frequency. This suggests that dividing the energy of each coefficient by an estimate from its neighborhood (Eq. (2)) may reduce the relations between the samples of the result. Note that the psychophysically-inspired neighborhood (bottom-left subplot

in Fig. 1, or Eqs. (3) and (4)) also increases with frequency as the statistical interaction neighborhood in Fig. 2.

In order to quantify the problems of linear representations and the potential advantages of the proposed normalized representation, we compared four representations (raw pixels, local DCT, local PCA, and the proposed normalized DCT) using four different measures of dependency (standard covariance, amplitude covariance, mutual information, and perceptual correlation). The results are given in table I. These measures were estimated from a training set consisting of 57344 blocks of size $16 \times 16$ taken from the Van Hateren database of calibrated natural images [49]. Each of the correlation measures (whether statistical or perceptual) are computed for all pairs of coefficients, thus forming a generic dependency matrix, $M$ (covariance, amplitude covariance and perceptual metric). The scalar measures shown in table I are computed by comparing the magnitude of the off-diagonal elements with the magnitude of the diagonal elements [2],

$$\eta = \frac{\sum_{i \neq j} |M_{ij}|}{\sum_i |M_{ii}|}. \tag{8}$$

The results in table I are consistent with our hypothesis regarding normalization. The first row of the table shows the interaction measure on the standard covariance, $\eta_s$. For this measure, the local-PCA representation, which is chosen to diagonalize the covariance matrix, achieves the best result. The local-DCT is known to provide a good fixed-basis approximation of local-PCA [2, 48], and performs quite well in comparison to the pixel basis. Surprisingly, the normalized representation is seen to be better than the local-DCT basis.

The second row of table I shows $\eta_{|s|}$, the interaction measure for higher-order statistics, as represented by covariance of coefficient amplitudes. This measure clearly indicates that the linear transforms do not succeed in removing these interactions, and thus do not lead to statistical independence. On the other hand, we see that the normalization representation greatly reduces these higher-order interactions.

The third row of table I provides a mutual information measure of the statistical independence of the coefficient amplitudes. The mutual information of a set of variables, $c_1, \cdots, c_n$, is defined as the Kullback-Leibler distance between their joint PDF and the product of their marginals, and it can be computed from the marginal entropies, $H(c_i)$, and joint entropy, $H(c_1, \cdots, c_n)$, of the variables [50]:

$$I(c_1, \cdots, c_n) = \sum_{i=1}^{n} H(c_i) - H(c_1, \cdots, c_n). \tag{9}$$

$I(c_1, \cdots, c_n)$ can be interpreted as the average number of bits that are lost when encoding the variables assuming they are independent. As the entropy of the coefficients, $H(c_i)$, may be quite different in each domain, we compute the relative mutual information, i.e., the *proportion of bits* that are lost when using a coder that assumes independence:

$$I_r(c_1, \cdots, c_n) = \frac{\frac{1}{(n-1)} I(c_1, \cdots, c_n)}{\frac{1}{n} \sum_{i=1}^{n} H(c_i)}. \tag{10}$$

Note that $I_r = 1$ when the $c_i$ are fully redundant (e.g., identical) and $I_r = 0$ when they are independent.

TABLE I
STATISTICAL AND PERCEPTUAL INTERACTION MEASURES FOR DIFFERENT REPRESENTATIONS.

| | *pixels* | *local-DCT* | *local-PCA* | *normalized-DCT* |
|---|---|---|---|---|
| $\eta_s$ | 158.3 | 7.2 | 0.0 | 0.8 |
| $\eta_{\|s\|}$ | 158.3 | 21.8 | 16.9 | 1.2 |
| $\eta_p$ | 47.6 | 1.4 | 12.1 | 0.0 |
| $I_r$ | 0.69 | 0.28 | 0.29 | 0.06 |

Because the estimation of information requires substantially more data than estimation of correlations, we restrict our relative mutual information calculation to a set of five coefficient amplitudes in each of the representations. In the spatial domain we considered the central coefficient and four neighbors around it (two horizontal and two vertical). In the PCA domain we took the first five coefficients after the first one (which approximately accounts for the average luminance). In the DCT and the DCT-normalized domains we considered the five AC coefficients of lower frequency. Histograms of 10 bins per dimension in the range $[0, \max(|c_i|)]$ were used to estimate the PDFs of the coefficient amplitudes. The $I_r$ results shown in table I are consistent with the reductions of the mutual information using divisive normalization reported elsewhere [10, 39, 51] and confirm the statistical benefits of the proposed representation.

Finally, the last row table I shows a perceptual interaction measure, $\eta_p$, computed from the perceptual metric matrix derived using Eq. (6) and assuming that the normalized domain is perceptually independent. Surprisingly, the local-PCA representation performs significantly worse than the local-DCT, even though it is optimized for second-order statistical independence. The results provide a quantitative measure of the claim made earlier in this section, that linear representations must necessarily have sub-optimal perceptual behavior.

Overall, we conclude that the superior statistical and perceptual properties of the divisive normalization representation, as compared with common linear representations, provide a justification for its use in transform coding.

## IV. INVERTING THE DIVISIVE NORMALIZATION TRANSFORM

In order to use a normalized representation directly in a transform coding application, we need to be able to invert the transformation. in this section, we study the *analytic* inversion problem and develop an efficient algorithm based on *series expansion*. A more general numerical inversion method (the *differential method*) was originally proposed in [11], and the advantages of this method were analyzed in [52]. However, the series expansion method proposed here is roughly three orders of magnitude faster than the differential method, and thus represents a better choice in practice. We also derive a general condition for the normalization to be invertible and verify that the psychophysically-derived normalization scheme used in this paper fulfills this condition.

Let $D_r$ and $D_\beta$ be diagonal matrices with the absolute value of the elements of $r$ and $\beta$ in the diagonal, then from Eq. (2)

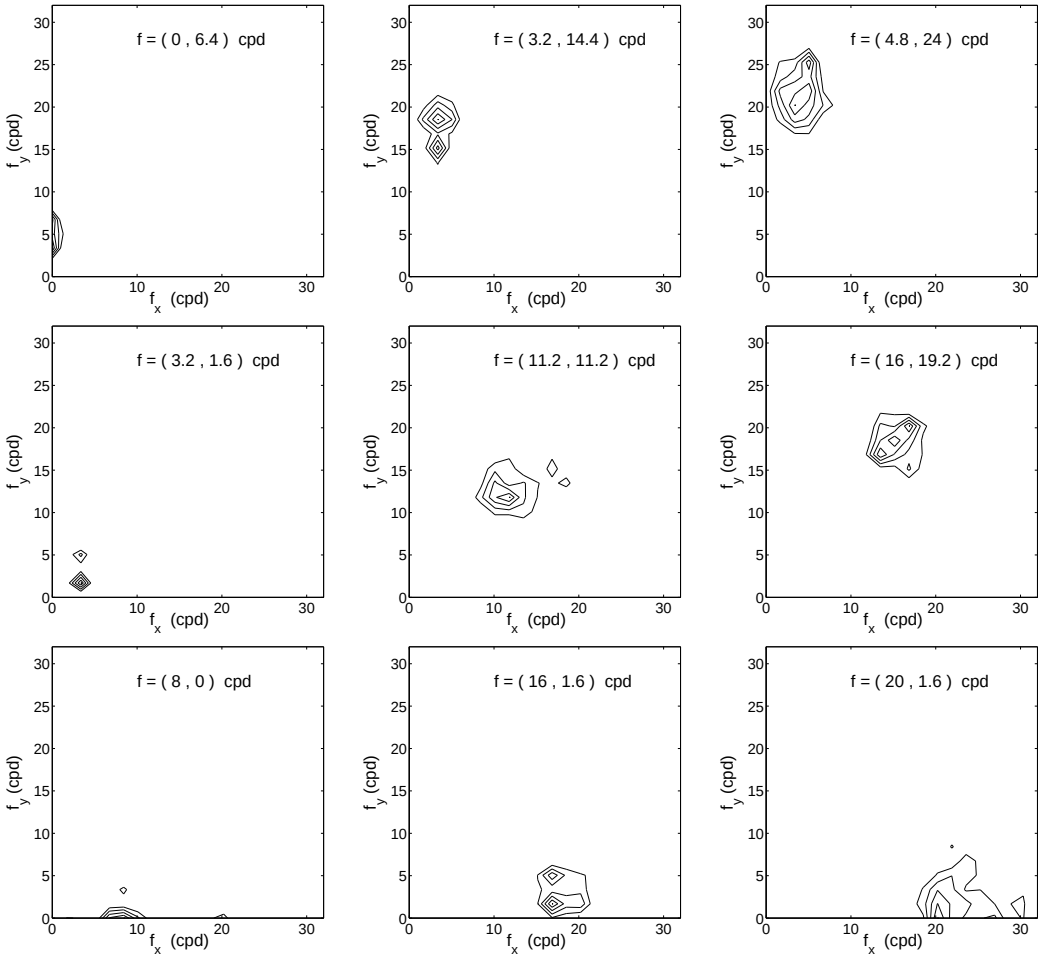



Fig. 2. Contour plots of the cross correlation of nine different local-DCT coefficient amplitudes with those at other frequencies and orientations. The nine representative coefficients were selected at three different frequencies (columns) and three orientations (rows). These interactions were measured in 57344 blocks of size $16 \times 16$ taken from the Van Hateren database of calibrated natural images [49].

it follows:

$$c = \mathrm{sgn}(r) \left( (I - D_r \cdot h)^{-1} \cdot D_\beta \cdot |r| \right)^{1/\gamma}. \qquad (11)$$

where, as in Eq. (2), the sign function $\mathrm{sgn}(\cdot)$, the absolute value $|\cdot|$, and the exponent $1/\gamma$ are applied in a element-by-element basis.

However, this analytic solution is not practical due to the computational cost of computing the inverse $(I - D_r \cdot h)^{-1}$. While computing the normalization transformation is efficient because the interactions between the coefficients of $c$ are local ($h$ is sparse), the inverse transformation suffers from global interactions between the coefficients of $r$ (i.e., the matrix $(I - D_r \cdot h)^{-1}$ is dense). Thus, direct calculation of Eq. (11) is costly even for moderate-sized images.

### 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 inversion or computation with dense matrices. The idea is using a series expansion of the inverse matrix in Eq. (11):

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

In that way we can compute the inverse up to a certain degree of approximation, $n$, taking a finite number of steps in the series:

$$\begin{aligned}
|c|^\gamma{}_{(1)} &= D_\beta \cdot |r| + (D_r \cdot h) \cdot D_\beta \cdot |r| \\
|c|^\gamma{}_{(2)} &= D_\beta \cdot |r| + (D_r \cdot h) \cdot D_\beta \cdot |r| + (D_r \cdot h)^2 \cdot D_\beta \cdot |r| \\
|c|^\gamma{}_{(3)} &= D_\beta \cdot |r| + (D_r \cdot h) \cdot D_\beta \cdot |r| + (D_r \cdot h)^2 \cdot D_\beta \cdot |r| + \\
&\quad (D_r \cdot h)^3 \cdot D_\beta \cdot |r| \\
&\vdots
\end{aligned}$$

A naive implementation would imply computing powers of $D_r \cdot h$ which is also a problem. However, it is possible to write the series approximation in a recursive fashion that only involves vector additions and matrix-on-vector multiplications:

$$\begin{aligned}
|c|^\gamma{}_{(0)} &= D_\beta \cdot |r| \\
|c|^\gamma{}_{(n)} &= D_\beta \cdot |r| + D_r \cdot h \cdot |c|^\gamma{}_{(n-1)}, \qquad (12)
\end{aligned}$$

Note that the matrices in Eq. (12) are sparse and thus the series may be computed efficiently.

### *B. General invertibility condition*

Despite the differences between the proposed inversion procedures (analytic and series expansion), the same condition has to hold to ensure the existence of the solution. This condition also applies for the previously reported differential method (see [52] for details).

Let $V$ and $\lambda$ be the eigenvector and eigenvalue matrix decomposition of $D_r \cdot h$:

$$D_r \cdot h = V \cdot \lambda \cdot V^T.$$

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

$$\lambda_{max} = max(\lambda_i) < 1. \tag{13}$$

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, $\lambda_i$, 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 examples in section IV-C), the matrix is likely to be ill-conditioned if the condition (13) doesn't hold.

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

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

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

### *C. Invertibility of psychophysically-inspired normalization*

We have empirically checked the invertibility of the normalization that uses psychophysically-inspired parameters for the local-DCT by computing the maximum eigenvalue of $D_r \cdot h$ over 25600 blocks randomly taken from the Van Hateren natural image data set [49]. Figure 3a shows the average eigenvalues spectrum and Fig. 3b the PDF of the maximum eigenvalue. In this experiment on a large natural data base the maximum eigenvalues are always far enough from 1. These results suggest that the normalization with these parameters will be invertible (see section IV-D), and it will remain invertible even if the responses $r$ undergo small distortions such as quantization (see section V-B).

### *D. Convergence rates*

In this section we analyze the convergence of the proposed inversion procedure. In the experiments shown here, we used the psychophysically-inspired parameters of section III and the local-DCT. Of course, such a simple (small size) transform does not really require iterative techniques because the analytical inverse is generally affordable.

It is possible to derive an analytic description for the convergence of the series expansion method. It turns out that the convergence is faster for a smaller $\lambda_{max}$. Consider that the error vector at the step $n$ of the approximation,

$$e_{(n)} = |c|^\gamma - |c|^\gamma_{(n)},$$

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

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

Then, taking the $|\cdot|_\infty$ norm as a measure of the error, we have that the error at each step is:

$$\epsilon_{(n)} = |e_{(n)}|_\infty = max(e_{(n)\,i}) \propto$$
$$\sum_{k=0}^{\infty} {\lambda_{max}}^{(n+k+1)} = \lambda_{max}^n \cdot \left( \frac{\lambda_{max}}{1 - \lambda_{max}} \right). \tag{14}$$

Figure 4 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 (blocks) with different $\lambda_{max}$. From Eq. (14) 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 Fig. 4 shows the result of local-DCT blocks, but the same behavior is obtained in the wavelet case [36].

## V. IMAGE CODING APPLICATION

Given the inversion results of the previous section, we can now consider the development of an image compression procedure based on a divisive normalization representation. Specifically, we propose to encode images using scalar quantization and entropy coding of a set of normalized local-frequency coefficients. The decoding procedure is then the natural one: first recover the quantized coefficients from the entropy-coded bitstream, then invert the normalization procedure, and finally invert the linear transform. In order to do this, we must first describe the quantizer design, and then verify the robustness of the invertibility condition in the presence of quantization errors and progressive coding.

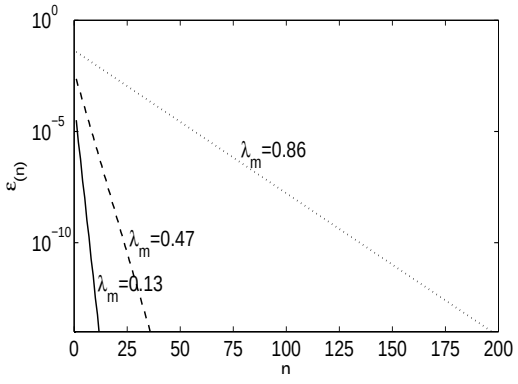



Fig. 4. Error of the series expansion method as a function of the number of terms in the series. The different lines represent the error obtained when inverting different images with different values of $\lambda_{max}$.

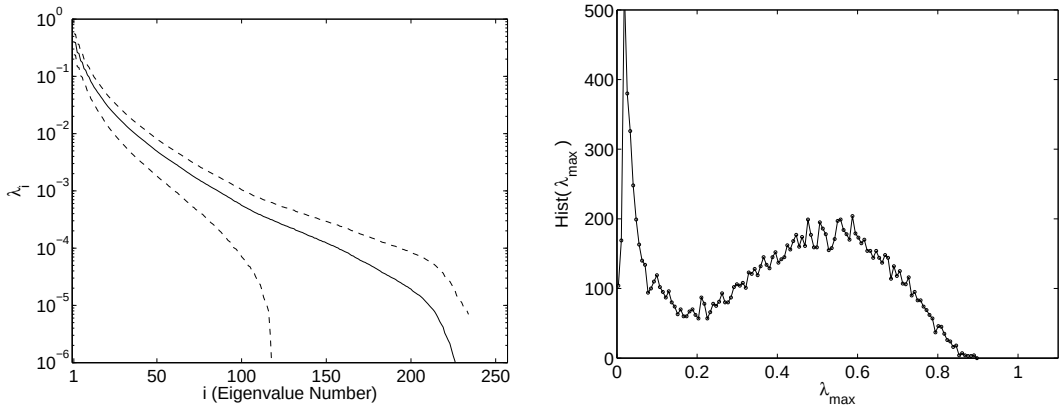



Fig. 3. Behavior of $D_r \cdot h$ for a set of 25600 blocks taken from the Van Hateren data base [49]. (a) Average eigenvalues spectrum. Dashed lines represent the standard deviation. (b) PDF of $\lambda_{\max}$

### A. Quantizer design

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 [3, 53] or some modification of them including the perceptual metric [17, 33–35]. However, it has been suggested that constraining the Maximum Perceptual Error (MPE) may be better than minimizing its average [33–35, 54]. This is because the important issue is not minimizing the average error across the regions but minimizing the annoyance in every quantized 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 quantization and *uniform* bit allocation. Of course, the expression of this quantizer turns out to be non-uniform in the linear transform domain (local-DCT or wavelets).

The difference between the approaches that implicitly followed the MPE idea [17, 22, 23, 31–35] is the accuracy of the perception model which is used to propose the perceptually Euclidean domain before the uniform quantization:

- JPEG [22] (and MPEG [23]) assume the linear CSF model [40]. This implies a fixed diagonal metric matrix in the DCT domain. This equivalence has been shown in [33–35].
- The algorithms of Daly [31], Watson [32] or Malo et al. [33–35], assume a point-wise non-linear model [37, 55]. This implies an input-dependent diagonal metric in the DCT domain.
- The algorithm of Epifanio et al. [17] uses the current non-linear model [25, 29, 38], i.e. it uses a non-diagonal metric. However, they use an average (input-independent) metric in the linear domain in order to avoid the inversion problem and to allow a linear simultaneous diagonalization of $\Gamma$ and $W$. It has to be stressed that this algorithm explicitly takes into account the image statistics using a local-PCA instead of a local-DCT.
- The proposed approach uses the current non-linear model [25, 29, 38] in the proper way: i.e. using the non-linear normalized representation and inverting it after the quantization. This means assuming an input-dependent and non-diagonal perceptual metric in the linear domain.

### B. Robustness of the invertibility under quantization

Figure 5 shows the effect of the quantization step (number of bits per coefficient) on $\lambda_{max}$. These results capture the evolution of the maximum eigenvalue of 100 256×256 images (25600 blocks) from the Van Hateren data base [49] when compressing them in the range $[0.02, 1.2]$ bits/pixel. For higher bit rates (over 1.5 bits/pix) the maximum eigenvalue remains stable and equal to its value in the original signal. For lower bit rates (as shown in the figure) $\lambda_{max}$ oscillates, but (for every block of these 100 representative images) always lies in the region that allows invertibility. At extremely low bit rates $\lambda_{max}$ tends to zero because the quantization step is so coarse that most of the coefficients in the quantized vector $\hat{r}$ are set to zero, inducing a reduction in the eigenvalues of $D_{\hat{r}} \cdot h$. This result suggests that the proposed normalized representation is invertible regardless of the bit rate. Thus, the coarseness of the quantization may be freely chosen to satisfy a distortion requirement.

Once we have shown that the quantization does not critically

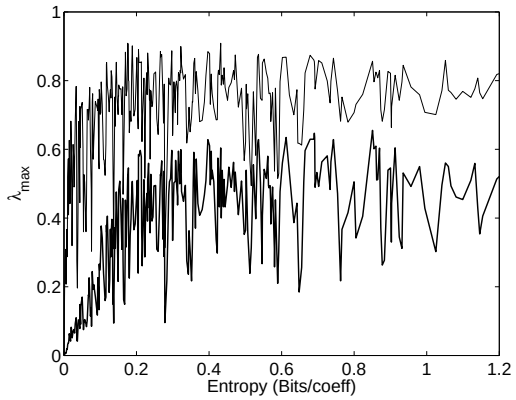



Fig. 5. Effect of quantization on $\lambda_{max}$. The thick line represents the average of $\lambda_{max}$ over the blocks of each image. The thin line shows the behavior of the maximum $\lambda_{max}$ in each image (worst case blocks).

affect the invertibility condition, another practical issue is robustness to progressive transmission, which is a convenient feature in any encoding scheme. Progressive coding refers to the ability to reconstruct different (approximated) versions of a fully decoded image from a subset of the bits of the compressed image. The proposed representation will be useful for progressive coding if taking a subset of the coefficients of an encoded image reduces the corresponding $\lambda_{max}$.

Figure 6 shows a representative example that illustrates the effect of progressive coding on $\lambda_{max}$. Figure 6 shows the evolution of the values of $\lambda_{max}$ for the blocks of the image Barbara as different proportions of coefficients are received by the decoder. In this example the original image was represented in the proposed domain and compressed to 0.58 bits/pix. Then, different reconstructions of the image are obtained when 4, 8, 12, 16 and 32 quantized coefficients (of increasing frequency) per block are received at the decoder. The values of $\lambda_{max}$ for the different subsets of coefficients are compared in each case with the values of $\lambda_{max}$ for the whole set of quantized coefficients. The trend illustrated by this example was also found in all the analyzed images. Using a small subset of coefficients substantially reduces $\lambda_{max}$. From this situation (coarse approximation), as more coefficients are sent to the decoder, the corresponding values of $\lambda_{max}$ progressively increase and tend to the values obtained using the whole set of coefficients. As the maximum eigenvalues of the incomplete signals are always below the corresponding values for the complete signal, the approximate image can always be reconstructed from the incomplete and quantized normalized representation.

### *C. Coding results*

In this section we compare the results of different MPE transform coding schemes described above: JPEG [22], the algorithm of Malo et al. [33–35] (which is similar to the algorithms of Daly [31] and Watson [32]), the algorithm of Epifanio et al. [17], and the proposed algorithm.

Figure 7 shows the average rate-distortion performance of the algorithms when coding five standard images in the range $[0.18 - 0.58]$ bits/pixel (Barbara, Boats, Lena, Einstein and Peppers). The distortion was computed using a standard objective distortion measure: the Peak-to-peak Signal-to-Noise Ratio (PSNR), defined as $10\log_{10}(255^2/\sigma_e^2)$, with $\sigma_e^2$ the error variance. The rate was computed using a standard zero-order entropy coding of the quantized coefficients.

The subjective performance of the algorithms can be seen in Figs. 8 and 9. They show some representative examples of the results: Barbara and Boats at 0.18 bits/pix. Note that these bit rates are substantially smaller than the usual bit rate recommended in the JPEG standard (between 0.5 and 1.0 bits/pix for achromatic images). These choices were made to ensure that the compression artifacts are substantially larger than those introduced by the journal printing process, thus allowing the reader to easily compare the visual quality of the algorithms. In the laboratory, we find that visual comparison of images at higher bit rates leads to analogous results. A sampling frequency of 64 samples/degree was assumed in the computations, so the viewing distance should be adjusted so that the angular extent of the (256×256) images is 4 degrees.

The JPEG results (Figs. 8b and 9b) exhibit over-smoothed areas because the width of the bit allocation function (the

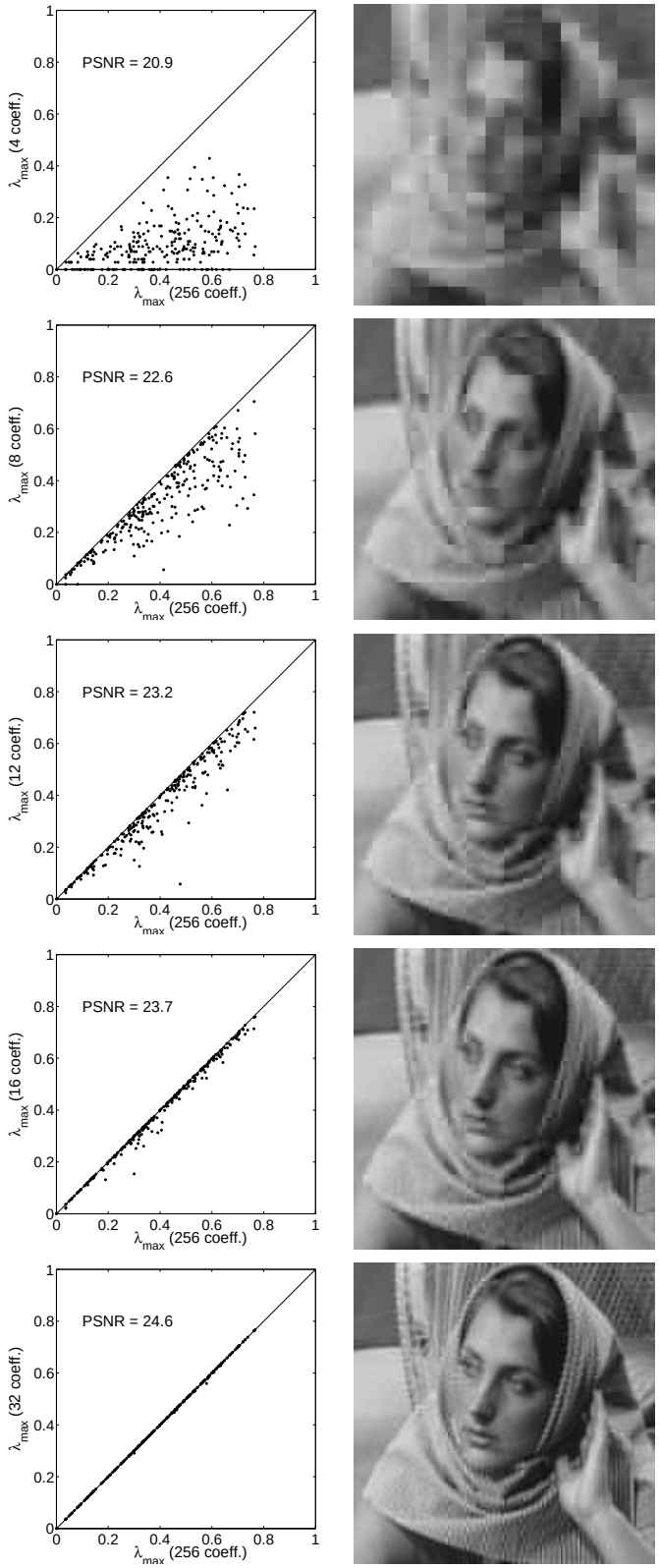



Fig. 6. Example of the evolution of $\lambda_{max}$, the reconstructed image and the PSNR in progressive coding.

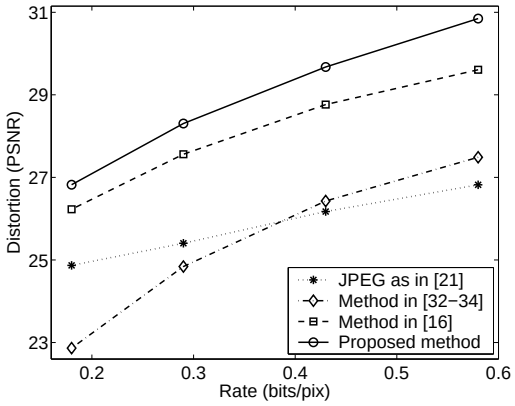



Fig. 7. Rate-distortion performance of the algorithms over 5 standard images (Barbara, Boats, Lena, Einstein and Peppers). JPEG [22] MPE quantizer with fixed diagonal metric (stars, dotted line), MPE quantizer using a point-wise non-linearity, i.e. adaptive diagonal metric [33–35] (diamonds, dash-dot line), MPE quantizer using a fixed non-diagonal metric [17] (squares, dashed line), the proposed approach: MPE quantizer in a normalized domain, i.e. adaptive non-diagonal metric (circles, solid line).

CSF) is too narrow. Therefore, high frequency textures are lost in the decoded images. As reported in the literature [32–34] the use of a point-wise non-linearity certainly preserves more high frequency details giving a better performance than JPEG at bit rates higher than 0.4 bits/pix (see the rate-distortion curves in Fig. 7). However, for very low bit rates the blocking effect is noticeably increased (Figs. 8c and 9c). The use of a simultaneously decorrelated linear domain (fixed but non-diagonal perceptual metric) improves the results but still adds high-frequency noise that is especially visible in the smooth regions (Figs. 8d and 9d). This effect comes from the PCA part of the linear decorrelating transform. The uniform quantization in the proposed normalized domain (Figs. 8e and 9e, and solid line in Fig. 7) gives rise to the best subjective results at every compression ratio in the analyzed range.

On the other hand, the intrinsic statistical power of the normalized representation is confirmed by the quality of the numerical (PSNR) results, as illustrated in Fig. 7. Note that the proposed representation increases the compression ratio by a factor of three or more with regard to the JPEG standard at the same PSNR level.

## VI. Conclusions

We have proposed the direct use of divisive normalization in transform coding. This nonlinear augmentation of a traditional linear transform leads to a substantial reduction in the both the perceptual and statistical dependencies between the coefficients of the representation. The combination of these two improvements implies that subsequent scalar processing and encoding of the coefficients can be nearly optimal in terms of both bitrate and perceptual distortion

We have studied the analytic invertibility of the divisive normalization representation, and proposed an efficient algorithm based on series expansion. When using a moderate size block-transform the analytical inversion is computationally affordable, but when using a wavelet basis, the series expansion inversion is the better choice. We also derived the general condition for the normalization to be invertible and showed that the proposed psychophysically-derived normalization is invertible. The empirical results on a large natural image collection suggest that quantization does not generally interfere with invertibility. However, it is still possible that inversion could fail on some particular images at some levels of quantization. In these cases the invertibility condition is a practical tool to detect this problem and solve it by slightly adjusting the bit rate.

Finally, image coding results suggest that a straightforward uniform quantization of the normalized coefficients using the psychophysically-inspired parameters is a simple and promising alternative to the current transform coding techniques that use perceptual information in the image representation and quantizer design. These results show that removing or reducing the (statistical and perceptual) dependence in linear transforms makes a big difference in the quality (or bit rate) of the reconstructed images.

The ability of the proposed representation to reduce the statistical dependence among the coefficients may alleviate the need for more sophisticated methods to extract any residual statistical relationships amongst the linear transform coefficients. Nevertheless, the results reported here could be improved by trying to exploit the statistical relations that may remain in the non-linear representation. However, it has to be stressed that the current techniques that exploit the redundancies in transform domains [8–10] should be substantially changed as the statistical nature of the signal in the non-linear representation is different [39].

Future work should consider alternative methods of estimating the parameters of the normalization (e.g., the statistical approach in [38]) which may improve the statistical benefits of the representation while retaining its perceptual properties. This effort is related to the development of more accurate statistical models for natural images. Finally, the properties of the proposed normalized representation may be useful in other image processing problems (e.g. denoising or texture analysis and synthesis) where both the perceptual and statistical properties of the coefficients are of fundamental importance.

## Acknowledgments

JM thanks the co-authors for their patience while he was slowly writing the manuscript. You just have to consider the fact that Fig. 7 represents the evolution of 15 years of perception based image coding. Therefore, one year delay is not that much.

## References

[1] J. Gutiérrez, F. Ferri, and J. Malo, "Regularization operators for natural images based on non-linear perception models," *To appear in: IEEE Transactions on Image Processing*, 2005.

[2] R. Clarke, *Transform Coding of Images*. New York: Academic Press, 1985.

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

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

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

[6] A. Bell and T. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vision Research*, vol. 37, no. 23, pp. 3327–3338, 1997. [Online]. Available: citeseer.nj.nec.com/bell97independent.html

[7] E. Simoncelli, "Statistical models for images: Compression, restoration and synthesis." in *31st Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA.*, 1997.

[8] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans Sig Proc*, vol. 41, no. 12, pp. 3445–3462, December 1993.

[9] A. Said and W. A. Pearlman, "An image multiresolution representation for lossless and lossy image compression," *IEEE Transactions on Image Processing*, vol. 5, no. 9, pp. 1303–1310, 1996.

[10] R. Buccigrossi and E. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 8, no. 12, pp. 1688–1701, 1999.

[11] J.Malo, R.Navarro, I.Epifanio, F.Ferri, and J.M.Artigas, "Non-linear invertible representation for joint statistical and perceptual feature representation," *Lect. Not. Comp. Sci.*, vol. 1876, pp. 658–667, 2000.

[12] D. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," *Physical Review Letters*, vol. 73, no. 6, pp. 814–817, 1994.

[13] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.*, vol. 46, pp. 886–902, April 1998.

[14] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Adv. Neural Information Processing Systems (NIPS*99)*, S. A. Solla, T. K. Leen, and K.-R. Müller, Eds., vol. 12. Cambridge, MA: MIT Press, May 2000, pp. 855–861.

[15] H. Choi and R. Baraniuk, "Multiscale image segmentation using wavelet-domain hidden markov models," *IEEE Trans. Image Proc.*, vol. 10, no. 9, Sep 2001.

[16] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, "Image denoising using a scale mixture of Gaussians in the wavelet domain," *IEEE Trans Image Processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.

[17] 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, pp. 1799–1811, 2003.

[18] P. Teo and D. Heeger, "Perceptual image distortion," *Proc. of the First IEEE Intl. Conf. Im. Proc.*, vol. 2, pp. 982–986, 1994.

[19] J. Malo, A. Pons, and J. 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, 1997.

[20] B. Girod, "What's wrong with mean-squared error," in *Digital Images and Human Vision*, A. B. Watson, Ed. the MIT press, 1993, pp. 207–220.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Perceptual image quality assessment: From error visibility to structural similarity," *IEEE Trans Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

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

[23] D. LeGall, "MPEG: A video compression standard for multimedia applications," *Communications of the ACM*, vol. 34, no. 4, pp. 47–58, 1991.

[24] D. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Boston: Kluwer Academic Publishers, 2001.

[25] A. Watson and J. Solomon, "A model of visual contrast gain control and pattern masking," *Journal of the Optical Society of America A*, vol. 14, pp. 2379–2391, 1997.

[26] J. Foley, "Human luminance pattern mechanisms: Masking experiments require a new model," *Journal of the Optical Society of America A*, vol. 11, no. 6, pp. 1710–1719, 1994.

[27] W. S. Geisler and D. G. Albrecht, "Cortical neurons: Isolation of contrast gain control," *Vision Research*, vol. 8, pp. 1409–1410, 1992.

[28] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual Neuroscience*, vol. 9, pp. 181–198, 1992.

[29] J. R. Cavanaugh, W. Bair, and J. A. Movshon, "Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons," *J Neurophysiology*, vol. 88, no. 5, pp. 2547–2556, November 2002.

[30] B. O. E.P. Simoncelli, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, pp. 1193–1216, 2001.

[31] S. Daly, "Application of a noise-adaptive Contrast Sensitivity Function to image data compression," *Optical Engineering*, vol. 29, no. 8, pp. 977–987, 1990.

[32] A. Watson, "DCT quantization matrices visually optimized for individual images," in *Human Vision, Visual Processing and Digital Display IV*, B. Rogowitz, Ed., vol. 1913, 1993.

[33] J. Malo, F. Ferri, J. Albert, and J. Soret, "Comparison of perceptually uniform quantization with average error minimization in image transform coding," *Electronics Letters*, vol. 35, no. 13, pp. 1067–1068, 1999.

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

[35] 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 Transactions on Image Processing*, vol. 10, no. 10, pp. 1411–1427, 2001.

[36] Y. Navarro, J. Rovira, J. Gutiérrez, and J. Malo, "Gain control for the chromatic channels in JPEG 2000," *To appear in: Proc. of the 10th Congress of the Intl. Colour Assoc.*, May 2005.

[37] G. Legge, "A power law for contrast discrimination," *Vision Research*, vol. 18, pp. 68–91, 1981.

[38] O. Schwartz and E. Simoncelli, "Natural signal statistics and sensory gain control," *Nature Neuroscience*, vol. 4, no. 8, pp. 819–825, 2001.

[39] J. Rovira, "Improving linear ICA with divisive normalization," MSc. Thesis, Dept. d'Òptica, Facultat de Física, Universitat de València, 2004.

[40] F. Campbell and J. Robson, "Application of Fourier analysis to the visibility of gratings," *Journal of Physiology*, vol. 197, pp. 551–566, 1968.

[41] E. Peli, "Contrast in complex images," *JOSA A*, vol. 7, pp. 2032–2040, 1990.

[42] M. Duval-Destin, M. Muschietti, and B. Torrésani, "Continuous wavelet decompositions: Multiresolution and contrast analysis," *SIAM J. Math. Anal.*, vol. 24, pp. 739–755, 1993.

[43] A. Pons, "Estudio de las funciones de respuesta al contraste del sistema visual," Ph.D. dissertation, Dpt. d'Òptica, Facultat de Física, Universitat de València, Julio 1997.

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

[45] B. Dubrovin, S. Novikov, and A. Fomenko, *Modern Geometry: Methods and Applications*. New York: Springer Verlag, 1982, ch. 3: *Algebraic Tensor Theory*.

[46] A. Ahumada, "Computational image quality metrics: A review," in *Intl. Symp. Dig. of Tech. Papers, Sta. Ana CA*, ser. Proceedings of the SID, J. Morreale, Ed., vol. 25, 1993, pp. 305–308.

[47] A. Pons, J. Malo, J. Artigas, and P. Capilla, "Image quality metric based on multidimensional contrast perception models," *Displays*, vol. 20, pp. 93–110, 1999.

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

[49] J. 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*, vol. 265, pp. 359–366, 1998.

[50] T. Cover and J. Tomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[51] R. Valerio and R. Navarro, "Optimal coding through divisive normalization models of V1 neurons," *Network: Comp. Neur. Syst.*, vol. 14, pp. 579–593, 2003.

[52] I. Epifanio and J. Malo, "Differential inversion of V1 non-linearities," Universitat de València," Tech. Rep., 2004.

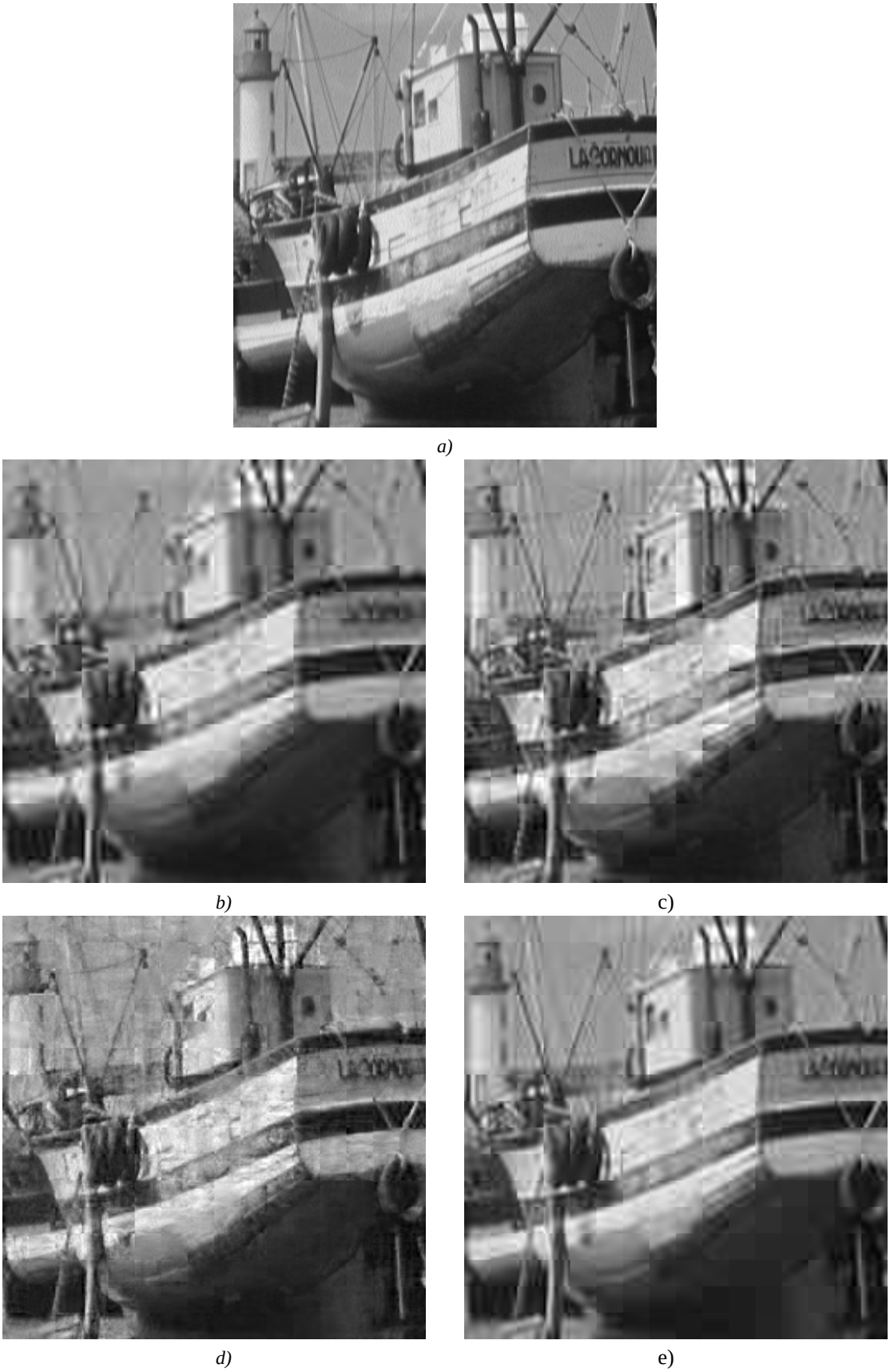
[53] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 127–135, 1982.

[54] G. Schuster and A. Katsaggelos, *Rate-Distortion Based Video Compression*. Boston: Kluwer Academic Publishers, 1997.

[55] A. Watson, "Efficiency of a model human image code," *Journal of Optical Society of America A*, vol. 4, no. 12, pp. 2401–2417, 1987.

*a)* *b)* c) *d)* e)

Fig. 8. Coding results (0.18 bits/pix). a) Original. b) JPEG [22] MPE quantizer with fixed diagonal metric (PSNR=23.7). c) MPE quantizer using a point-wise non-linearity (adaptive diagonal metric) [33–35] (PSNR=23.0). d) MPE quantizer using a fixed non-diagonal metric [17] (PSNR=24.3). e) The proposed approach: MPE quantizer in a normalized domain (adaptive non-diagonal metric) PSNR=26.5.



*a)* *b)* c) *d)* e)

Fig. 9. Coding results (0.18 bits/pix). a) Original. b) JPEG [22] MPE quantizer with fixed diagonal metric (PSNR=24.3). c) MPE quantizer using a point-wise non-linearity (adaptive diagonal metric) [33–35] (PSNR=23.7). d) MPE quantizer using a fixed non-diagonal metric [17] (PSNR=25.7). e) The proposed approach: MPE quantizer in a normalized domain (adaptive non-diagonal metric), PSNR=26.1.

**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 (A.B. Watson), and at the Lab of Computational Vision of the Center for Neural Science, New York University (E.P. Simoncelli). 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, la Bola de Cristal, and beauty in general...

**Irene Epifanio** was born in Valencia, Spain, in 1975. She graduated in Mathematics in 1997 and received the Ph.D. degree in Statistics in 2002, both from the Universitat de València, Valencia, Spain. In 1999 she joined the Computer Science Department, Universitat de Valencia. In October 2000, she joined the Department of Mathematics, Universitat Jaume I, Castello, Spain, where she is an Assistant Professor. Currently her research interests are focused on texture analysis and image compression.

**Rafael Navarro** received the MS and PhD degrees in Physics from the University of Zaragoza, Spain in 1979 and 1984, respectively. From 1985 to 1986 he was an optical and image processing engineer at the Instituto de Astrofísica de Canarias. In 1987 he joined the Instituto de Óptica "Daza de Valdés", Consejo Superior de Investigaciones Científicas, where, at present, is Professor of Research. Since 1988 he has headed the Imaging and Vision research group. In the period 1994-1999 he was associate director, and in 1999-2003 director of the Instituto de Óptica. He has been visiting researcher at the University of Rochester and at the University of California, Berkeley, and is member of the EOS, OSA, IEEE Signal Processing and ARVO. His research interests are Physiological Optics, Vision (human and artificial) and Image Processing, having contributed with about 65 papers in international SCI journals.

**Eero P. Simoncelli** received the B.S. degree in Physics in 1984 from Harvard University, studied applied mathematics at Cambridge University for a year and a half, and then received the M.S. degree in 1988 and the Ph.D. degree in 1993, both in Electrical Engineering from the Massachusetts Institute of Technology. He was an Assistant Professor in the Computer and Information Science department at the University of Pennsylvania from 1993 until 1996. He moved to New York University in September of 1996, where he is currently an Associate Professor in Neural Science and Mathematics. In August 2000, he became an Associate Investigator of the Howard Hughes Medical Institute, under their new program in Computational Biology. His research interests span a wide range of topics in the representation and analysis of visual images, in both machine and biological vision systems.