Perceptually and Statistically Decorrelated Features for Image Representation: Application to Transform Coding

Jesús Malo, Francesc Ferri[†]
Universitat de València
Dpt. d'Òptica, [†]Dpt. d'Informàtica
C/ Dr. Moliner 50, 46100 Burjassot, Spain
jesus.malo@uv.es

Rafael Navarro, Roberto Valerio
Instituto de Óptica (CSIC)
C/ Serrano 121, 28006 Madrid, Spain
r.navarro@io.csic.es

Abstract

Transform coding consists of a scalar quantization of the features of an image representation. These features should be independent enough to justify the scalar approach. The coefficients of the commonly used DCT representation still show some dependence that may reduce its efficiency. In this work, a perceptually inspired non-linear transform is used to map the DCT into a new representation that largely reduces the statistical and perceptual relations between the coefficients thus improving the compression performance[1].

1. Introduction

Independence between features is a desirable property of a signal representation in many image processing applications (such as indexing, fusion or transform coding) because it allows simple scalar data processing. In particular, in the transform quantization approach to image coding, the original image is represented in a feature space in order to simplify the quantizer design. Assuming that the transform removes the dependencies between coefficients, each feature is (scalarly) quantized according to its relevance [4].

The local DCT and wavelet transforms, are widely used in image coding due to both statistical and perceptual reasons. On one hand, they are used as an approximate *Principal Component Analysis* (PCA) that is supposed to remove the statistical correlations [2]. On the other hand, these transforms are similar to the first linear stage in human visual system processing. As a consequence, it is usually assumed that the distortion on one coefficient do not induce perceptual artifacts on other image components [6, 7, 10].

[1]This work has been partially supported by the projects CICYT-FEDER (TIC) 1FD-97-0279 and 1FD-97-1910, and CICYT (TIC) 98-677-C02

In spite of the good properties of these transforms, none of the above statistical and perceptual independence assumptions is completely true. First, linear PCA removes the second-order statistical relations only in the case of Gaussian statistics (which is not a valid assumption for natural images [1]). And second, the most recent human visual system (HVS) models include non-linear interactions between the linear transform coefficients [11], so the corresponding basis functions are not perceptually independent. These feature correlations may reduce the efficiency of the transforms used in image coding.

Some authors [9] suggest that the perceptual non-linear stage after the linear decomposition could be used by the HVS as a component independization transform. However, these non-linear interactions have not been used in image processing applications that require signal reconstruction because they are not analytically invertible.

In this work, a perceptually inspired non-linear transform [9] is used to map the local DCT into a new representation that largely reduces the statistical and perceptual dependence between the coefficients. A quasi-analytical method is proposed to invert this non-linear transform. This new domain does allow a strict scalar quantization of the coefficients for coding purposes. The compression performance in the proposed domain is compared to the usual DCT domain using several quantizer designs.

2. Statistical and Perceptual dependence between Features

The covariance matrix, Γ, is commonly used to evaluate the decorrelating efficiency of a feature extraction transform in the context of image coding [2]. However, there is no equivalent description of the perceptual interdependence between coefficients that could be used in the same way. In this section, an easily computable measure of the perceptual relations is presented: the *perceptual metric matrix*, W.

Any suitable feature extraction transform should simultaneously diagonalize both matrices Γ and W.

2.1. Covariance Matrix

The covariance matrix around a certain pattern, a_0, is given by the expected value of the outer product of the deviations from this point over a training set [2],

$$\Gamma(a_0) = \mathcal{E}\left[\Delta a \cdot \Delta a^T\right] \tag{1}$$

where $\Delta a = (a - a_0)$. It describes the shape of the (ellipsoidal) distribution of image samples around a_0. Non-zero off-diagonal elements indicate a deviation between the axis of the ellipsoid and the axis of the space, which means a strong statistical correlation between features.

2.2. Perceptual Metric Matrix

In a clustering problem, the data distribution is not the only parameter to be considered, but also the qualitative nature of the feature space. In image coding problems the perceptual geometry of the representation domain has to be taken into account. Given a certain image representation, such as a local DCT, the perceptual distortion induced in a pattern, a_0, due to a distortion Δa can be expressed as,

$$\begin{aligned} d(a_0, a_0 + \Delta a)^2 &= \Delta a^T \cdot W(a_0) \cdot \Delta a \\ &= \sum_i W_{ii} \Delta a_i^2 + 2 \sum_{i \neq j} W_{ij} \Delta a_i \Delta a_j \end{aligned} \tag{2}$$

where $W(a_0)$ is the perceptual metric at that point. The perceptual metric describes the shape of the (ellipsoidal) locus of perceptually equidistant patterns from a_0. The diagonal elements of the perceptual metric represent the contribution of each coefficient to the global distortion. Non-zero off-diagonal elements induce additional contributions to the distortion due to combinations of deviations in different dimensions, i.e. they represent *perceptual interactions* between features that modify the distortion perception.

It is important to know how the perceptual metric changes under mappings to evaluate its properties in different representation spaces. Invoking the usual invariance of the distance under changes of representation [3], it is easily obtained that, locally,

$$W' = (\nabla R^T)^{-1} \cdot W \cdot \nabla R^{-1} \tag{3}$$

2.3. Interaction measures

The covariance and perceptual metric describe ellipsoids, which when deviated from the axis of the space represent statistical or perceptual interactions between the features. The decorrelating efficiency of a feature extraction transform has been traditionally referred to the diagonal nature of the resulting covariance matrix [2]. The same concepts can be applied to the perceptual interaction descriptor, W, proposed here.

Given a (statistical or perceptual) matrix, M, that describe the second-order relations between features, a global *interaction measure* can be defined comparing the magnitude of the off-diagonal coefficients with the magnitude of the diagonal coefficients,

$$\eta = \frac{\sum_{i \neq j} |M_{ij}|}{\sum_i |M_{ii}|} \tag{4}$$

Any suitable feature extraction transform should simultaneously minimize the statistical interaction, η_s, and the perceptual interaction, η_p.

2.4. Perception models and DCT perceptual metric

The standard model of human low-level image analysis has two basic stages, in which the input image, A, is first transformed into a vector, a, in a local frequency domain (using a linear filter bank, T) and then it is non-linearly transformed into a response vector, r:

$$A \xrightarrow{T} a \xrightarrow{R} r \tag{5}$$

The perceptual transform, T, is similar to the class of transforms employed in image coding. The local DCT will be used here as a model of the perceptual linear transform T.

Different models of response, R, have been used to propose DCT quantizers. Simplified (linear) models of R are the foundation of the quantizer design in the JPEG standard. Non-linear models without interactions between coefficients have been applied in posterior improvements of the quantizer design [6, 7, 10]. The most recent response models (not yet applied in image coding) include interactions between coefficients of different frequencies and orientations [9, 11]. The response model that has been used here is basically the energy-normalization model of refs. [9, 11] plus a small linear term to regularize the gradient,

$$r_i = \alpha_i \frac{|a_i|^2}{\beta_i + (h * |a|^2)_i} + \frac{\alpha_i}{100}|a_i| \tag{6}$$

where the index, i, corresponds to spatial frequency. In this expression the energy of each transform coefficient is normalized by the convolution of the energy of the signal with a unit-volume Gaussian kernel. The Gaussian kernel, h, relates each coefficient with its neighbors in frequency and orientation. The width of the kernel increases with frequency because the length of the interaction is about one octave. The constants α and β, are chosen to fit the psychophysical response of the human viewer to periodic functions of frequency, i, and amplitude a_i.

Assuming the above T and R transforms, the explicit expression of the perceptual metric in the DCT domain can be obtained from eq. 3 and the geometric properties of the response domain.

The standard model of human pattern discrimination states that every component of the response vector has the same perceptual relevance and that there is no interaction between them [11]. In our formulation this implies an Euclidean (identity) perceptual metric in the response domain. Taking $W' = I$, the metric in the DCT domain is,

$$W(a) = \nabla R(a)^T \cdot \nabla R(a) \quad (7)$$

where the gradient is given by,

$$\nabla R(a)_{ij} = 2\alpha_i \left(\frac{|a_i|}{D(a)_i}\delta_{ij} - \frac{|a_i^2 \cdot a_j|}{D(a)_i^2} h_{ij} \right) + \frac{\alpha_i}{100}\delta_{ij} \quad (8)$$

and, $D(a)_i = \beta_i + (h * |a|^2)_i$, stands for the normalizing denominator. Note that the perceptual metric in the transform domain is *input-dependent* and *non-diagonal*.

2.5. Dependence measures in different domains

The measures η_s and η_p were computed in the spatial and the DCT domains to point out that additional post-DCT processing is required. This is because, despite the substantial reduction of both interactions the DCT does not remove none of them. Two significative post-DCT linear mappings have been explored to remark the unability of linear feature extraction transforms to simultaneously reduce both correlations. One of them is the transform that diagonalizes the covariance matrix (which is a standard PCA or KLT representation of the data and hence should give $\eta_s = 0$). The other is the linear transform that diagonalizes the perceptual metric in order to remove the subjective correlation.

The covariance matrix and the average perceptual metric were computed over a training set constituted by 1536 samples (16×16 blocks) obtained from six natural images. The first columns of the table 1 show the results for η_s and η_p in the spatial domain, in the DCT domain and in the additional representation spaces linearly obtained from the DCT.

These results show that when one tries a linear approach to remove one the correlations, the result is that the other increases. It is clear that any of transformed representations is better than pixels in the spatial domain, but strictly speaking, none of them is suitable to be used scalar-wise.

	x	DCT	*diag.*Γ	*diag.*W	Norm.
η_s	41.5	3.9	0.0	4.2	0.5
η_p	59.8	1.5	6.0	0.0	0.0

Table 1. Feature interaction measures in different representation domains.

The use of scalar quantizers on vector signals with correlated coefficients reduces the efficiency of the encoder [4]. Thus, a different representation space with a reduced correlation between features may improve the performance obtained in the DCT domain.

3. Component independization through non-linear DCT normalization

In this work we propose the use of the non-linear transform normalization of eq. 6 as a component independization mapping prior to scalar quantization in transform coding.

There are two reasons to use this transform. On one side, due to its nature, it completely removes the perceptual interaction because it gives rise to a perceptually Euclidean representation. On the other side, it has been reported [9] that such normalization can reduce the statistical relations between the outputs r_i.

The last column of the table 1 shows the interaction measures of the same training set in the non-linearly normalized DCT representation.

The statistical decorrelation efficiency of this transform can be explained in terms of a *divisive* DPCM. In predictive coding the value of one signal element is first predicted from the values of its neighbors, and then a decorrelated signal is obtained by *discounting the prediction*. In the commonly used DPCM the discount is linear: the prediction is subtracted from the input giving rise to an *error signal* which is highly decorrelated [4]. In the dominant term of eq. 6 the discount is divisive (non-linear). In this case, the energy of each transform coefficient is normalized by a function of the neighbor energies (by the convolution $\beta_i + (h * |a|^2)_i$). The convolution by the kernel h can be considered a linear estimation of the energy of each coefficient from the energies of its neighbors. The normalization by the prediction gives rise to a response signal in which the predictable part has been discounted.

4. Inversion of the DCT normalization

Despite the good statistical and perceptual decorrelating properties of the proposed representation, it could not be used in image coding applications without an efficient method to obtain the image from the quantized features.

Unfortunately, the normalization of eq. 6 is not analytically invertible because of the interaction kernel. As h is not diagonal, each response r_i is coupled with every transform coefficient a_j. Therefore, the inversion of R gives rise to a set of non-linear equations which can not be solved analytically. There are, of course, a number of numerical methods based on the iterative search of a solution, $\widehat{a}$, which minimizes some distance, $|r - R(\widehat{a})|$, but they may have slow convergence rates and be very sensitive to the initialization.

In this work, a *quasi-analytical* method to invert eq. 6 is presented. In spite of the non-invertible nature of R, if a linear piece-wise approximation is considered, the inverse function can be locally written as,

$$a = R^{-1}(r_0 + dr) = a_0 + \nabla R^{-1}(r_0) \cdot dr \qquad (9)$$

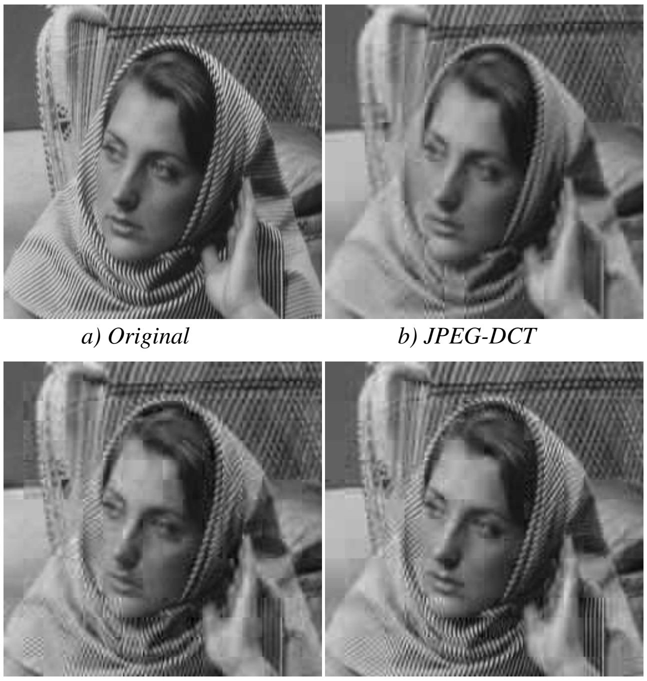
where the unknown gradient of the inverse function can be be related to the (known) gradient of the response: $\nabla R^{-1}(R(a_0)) = \nabla R(a_0)^{-1}$. This differential equation represents the local evolution of the inverse response. If it is integrable, it can be used to propagate the solution from any initial conditions, (r_0, a_0), up to the desired point r. The existence and uniqueness [5] of the solution of this *initial value problem* is ensured because the inverse of the gradient is always bounded. Note that the linear term has a negligible effect on the response ($0.01\alpha \ll \alpha$), but it avoids singularities in the inverse of the gradient.

The inverse is then obtained integrating the differential equation 9. Despite the differential equation is not analytically integrable, this method is essentially better than a purely numerical (search-based) method because the exact solution is guaranteed no matter the initial conditions. Experimental results [8] indicate that, for most responses, a simple four-order Runge-Kutta algorithm adequately converges in 7 or 8 steps.

5. Quantization results

In this section the same quantizer design criterion has been used in the DCT and in the independized domain to point out the advantages of the proposed representation. The design is based on the recently proposed *Maximum Perceptual Error* restriction (MPE) which includes JPEG quantization as a particular case and improves the classical perceptually weighted MSE approaches [6, 7]. Here, an Euclidean-uniform bit allocation quantizer in the normalized DCT domain is compared with two previously reported quantizers in the DCT domain: a JPEG-like quantizer (MPE with a linear response model) and a MPE quantizer with a non-interacting non-linear response [6, 7].

The euclidean MSE and the perceptual MSE (computed with the metric W) in the DCT approaches are about 40% greater than the equivalent distortions in the normalized DCT results. Figure 1 shows an example of the decoded Barbara image at the same compression ratio (0.27 bit/pix). It can be seen that the JPEG quantizer removes some significant high-frequency and high-contrast details giving a too smooth reconstruction. The non-linear DCT quantizer do preserve these details, but the blocking effect is increased. The proposed quantizer in the normalized domain gives rise to a better overall performance as can be seen in the pictures and in the distortion results.



a) Original *b) JPEG-DCT*

c) Non-linear MPE-DCT *d) Normalized DCT*

Figure 1. Barbara at 0.27 bits/pix.

References

[1] R. Buccigrossi and E. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans. Im. Proc.*, 8(12):1688–1701, 1999.

[2] R. Clarke. *Transform Coding of Images*. Academic Press, NY, 1985.

[3] B. Dubrovin, S. Novikov, and A. Fomenko. *Modern Geometry*. Springer Verlag, NY, 1982.

[4] A. Gersho and R. Gray. *Vector Quantization and Signal Compression*. Kluwer Acad. Publ., Boston, 1992.

[5] D. Logan. *Introduction to non-linear partial differential equations*. Wiley&Sons, NY, 1994.

[6] J. Malo, F. Ferri, J. Albert, J.Soret, and J. Artigas. The role of perceptual contrast non-linearities in image transform coding. *Im. Vis. Comp.*, 18(3):233–246, 2000.

[7] J. Malo, F. Ferri, J. Albert, and J. Soret. Comparison of perceptually uniform quantization with average error minimization in image transform coding. *Electr. Lett.*, 35(13):1067–1068, 1999.

[8] J. Malo, R. Navarro, I. Epifanio, F. Ferri, and J. Artigas. Non-linear invertible representation for joint statistical and perceptual feature decorrelation. *To appear in: Lect. Not. Comp. Sci., Proc. SPR+SSPR 2000, Springer Verlag*, 2000.

[9] E. Simoncelli and O. Schwartz. Modeling surround suppression in V1 neurons with a statistically-derived normalization model. In M. Kearns, editor, *Adv. in Neural Inf. Proc. Syst.*, volume 11. MIT Press, 1999.

[10] A. Watson. DCT quantization matrices visually optimized for individual images. *Proc. SPIE*, 1913, 1993.

[11] A. Watson and J. Solomon. A model of visual contrast gain control and pattern masking. *J. Opt. Soc. Am. A*, 14:2379–2391, 1997.