Non-linear Invertible Representation for Joint Statistical and Perceptual Feature Decorrelation

J. Malo[1], R. Navarro[3], I. Epifanio[2], F. Ferri[2], and J.M. Artigas[1]

[1] Dpt. d'Òptica, Universitat de València
[2] Dpt. d'Informàtica, Universitat de València
Av. Vicent Andrés Estellés S/N, 46100 Burjassot, València, Spain
[3] Instituto de Óptica (CSIC)
C/ Serrano 122, 28006 Madrid, Spain

Abstract. The aim of many image mappings is representing the signal in a basis of decorrelated features. Two fundamental aspects must be taken into account in the basis selection problem: data distribution and the qualitative meaning of the underlying space. The classical PCA techniques reduce the *statistical* correlation using the data distribution. However, in applications where human vision has to be taken into account, there are *perceptual* factors that make the feature space uneven, and additional interaction among the dimensions may arise.

In this work a common framework is presented to analyze the perceptual and statistical interactions among the coefficients of any representation. Using a recent non-linear perception model a set of input-dependent features is obtained which simultaneously remove the statistical and perceptual correlations between coefficients. A fast method to invert this representation is also presented, so no input-dependent transform has to be storaged. The decorrelating power of the proposed representation suggests that it is a promising alternative to the linear transforms used in image coding, fusion or retrieval applications[1].

1 Introduction

Independence among the features is recognized as an intrinsic advantage of a given signal representation because it allows simple scalar data processing and a better qualitative interpretation of the feature vector [1, 2]. This is why the aim of most feature extraction transforms is to find out a complete set (a basis) of independent features. Two main factors should determine the basis selection problem: the data distribution and the qualitative (geometric) properties of the underlying space. The basis functions should not only reflect the principal axis of the trainning set but also the eventual anisotropies of the space.

This is particularly important in applications involving natural imagery or texture description, such as indexing and retrieval [3], fusion [4], or transform

[1] The authors wish to thank E.P. Simoncelli for his fruitful comments. This work has been partially supported by the CICYT-FEDER (TIC) projects 1FD97-0279 and 1FD97-1910 and CICYT (TIC) 98-677-C02

coding [2]. In these cases, in addition to the data distribution, it is usually necessary to take into account the properties of Human Visual System (HVS): not every scale, texture or color component has the same relevance for the HVS, and undesired perceptual interactions among the coefficients may arise if they are scalarly processed. Therefore, in many applications the concept of independence of image coefficients has not only a statistical meaning, but it may also be related to the intrinsic (perceptual) geometry of the space. On the other hand, the HVS has developed efficient representations to deal with natural imagery [5–9], so the knowledge of the geometry of the low-level representation of a general-purpose biological vision system is of theoretical interest for image processing.

Wavelet and local DCT transforms, are widely used in many applications due to both statistical and perceptual reasons. On one hand, they are used as an approximate fixed-basis *Principal Component Analysis* (PCA). On the other hand, these transforms are similar to the first linear stage in HVS processing. However, the statistical and perceptual decorrelation obtained with these transforms is not complete.

Recently developed perception models that include non-linear interactions between the coefficents of a wavelet-like representation [10–12], can show interesting statistical decorrelation properties [9], but they cannot be used in image processing applications because they are not analytically invertible.

In this work the basis selection problem is analyzed from both the statistical and the perceptual points of view. Here the perceptual metric matrix is used as a complement of the covariance matrix to evaluate the statistical and perceptual interactions between the features under a common framework. Also, a fast method to invert the most recent perceptual representation [9–12] is developed and tested. It is shown that excellent decorrelation results are obtained from both statistical and perceptual points of view just taking into account the perceptual geometry of the wavelet-like feature space derived from this perception model. In this context, the decorrelating power of this representation is compared with fixed linear transforms and (unpractical) PCA-like methods that require the storage of ad-hoc basis functions.

2 Aim of the Feature Extraction Transform

Matrices of Second Order Relations. The *statistical deviations* from an image a_0 in a certain feature space are described by the covariance matrix, Γ:

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

Assuming a L^2 norm [11], the *perceptual deviation* from a_0 due to a distortion Δa is determined by the perceptual metric, W, of the domain at that point,

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

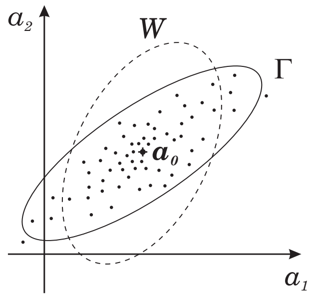

Fig. 1. Ellipsoids describing the data distribution and the space geometry around a_0.

Associated non-aligned ellipsoids. The covariance and the perceptual metric matrices are cuadratic forms that describe two different interesting ellipsoids.

On one hand, Γ describes the shape of the distribution of image samples around a_0. Non-zero off-diagonal elements in Γ indicate a deviation between the axis of the data ellipsoid and the axis of the space. This deviation implies a *statistical correlation* between features in the training set. On the other hand, W describes the shape of the (ellipsoidal) locus of perceptually equidistant patterns from a_0. The diagonal elements of W represent the contribution of each coefficient to the perceived distortion (eq. 2). 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 perceived distortion. This is a convenient way to represent what is commonly referred to as *masking*: a distortion in one coefficient may mask the subjective distortion in another one.

In the most general case these two ellipsoids are not aligned, so their eigenaxis, and the corresponding PCA-like basis functions, are not the same.

Measuring the Statistical and Perceptual Relations Among Features. The decorrelating efficiency of a feature extraction transform has been traditionally referred to the diagonal nature of the resulting covariance matrix. As the non-diagonal elements in W represent the 2nd-order perceptual interactions between the dimensions of the feature space, here we propose to evaluate the transforms from the perceptual point of view applying to W the same measures that have been used for Γ in the context of transform coding [2].

In this way, given a matrix, M, that describe the (statistical or perceptual) relations between the features, a scalar measure (the statistical interaction, η_s, or the perceptual interaction, η_p), can be defined comparing the magnitude of the off-diagonal coefficients with the magnitude of the diagonal coefficients,

$$\eta = \frac{\sum_{i \neq j} |M_{ij}|}{\sum_i |M_{ii}|} \tag{3}$$

Aim of the Feature Extraction Transform. In order to minimize the final correlations from both statistical and perceptual points of view, the transform should find out the eigenaxis of both ellipsoids, i.e., it should simultaneously diagonalize, Γ and W, or simultaneously minimize η_s, and η_p.

Given a perceptual matrix (or a certain input a_0) and using simple linear algebra it can be obtained the linear transform that simultaneously removes both correlations [13]. However, due to the highly point-dependent nature of W (see section 3.2), a different linear transform would be necessary for each possible input, which is not a practical solution.

In this work we will take a different approach: we will use the current non-linear perceptual model [9–12] to map a local DCT representation in a perceptually Euclidean space. In this way it is obtained $\mu_p = 0$ for every input. Beyond the obvious perceptual decorrelation, we will show that this non-linear transform has also statistical interest: due to its particular structure, it also removes the residual statistical correlations that remain in the DCT, strongly reducing μ_s. In this way, both measures, μ_s and μ_p, are minimized by a single adaptive transform which can be inverted without the storage of ad-hoc basis functions.

3 Visual Models and associated Perceptual Geometry

3.1 Metric and Visual Response

The standard model of human low-level image analysis has two basic stages, in which the input image, A, (defined in the *spatial domain*) is first transformed into a vector, $a = T \cdot A$, in a local frequency domain (the *transform domain*) using a linear filter bank, T, and then a set of mechanisms respond to each coefficient of the transformed signal giving an output, $r = R(a)$, which is the image representation in the *response domain*:

$$A \xrightarrow{T} a \xrightarrow{R} r \tag{4}$$

It is well established that the first linear perceptual transform, T, is similar to the class of wavelet-like transforms employed in many image analysis applications. This is not a casual result, because the low-level algorithms used by the HVS should be mainly determined by the statistics of natural images [5–9], and, as a result, linear PCA-like solutions have been developed in the low-level HVS.

Not all the basis functions of the transform T are equally perceived so additional processing, R, is included to explain these non-homogeneities. The HVS models assume that all the components of the r vector are equally important and there is no perceptual interaction between them [9, 11, 12] (i.e. the response domain is Euclidean), so the (perceptual) geometry of the transform domain (and also of the spatial domain) must depend on the nature of the response.

Given a response model, R, an explicit expression for the perceptual metric in any representation space can be obtained. The change of the elements of a tensor under a coordinate mapping depend on the jacobian of the transform [14]. Applying the general expressions for tensor transformation to our case, eq. 4, and assuming that the metric in r is the identity, the following metrics, W and w, in the transform and the spatial domains are obtained [15],

$$\begin{aligned} W(a) &= \nabla R(a)^T \cdot \nabla R(a) \\ w(A) &= T^T \cdot \nabla R(T{\cdot}A)^T \cdot \nabla R(T{\cdot}A) \cdot T \end{aligned} \tag{5}$$

where ∇R is the gradient (or jacobian matrix) of the non-linear response.

With these expressions and a particular perception model, a (T, R) pair, the interaction measure, μ_p, can be computed in any representation domain to evaluate the perceptual decorrelation power of any feature extraction transform.

3.2 Non-Linear Energy Normalization Model

The current models for R assume that after the application of the linear filter bank, T, the energy of each transform coefficient is normalized by a weighted sum of the energy of its neighbours. The dependence with the neighbour coefficients is given by the convolution with an interaction kernel h [9, 11, 12],

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

Figure 2 shows the parameters of this *non-linear energy normalization model* and an example of the response for some basis functions of different frequencies.

The parameters α_i, fig. 2.a, define a band-pass function that modulates the strenght of the response for each coefficient i. The parameters β_i, fig. 2.b, determine the point of maximum slope in each response. The values of α and β have been fitted to reproduce amplitude discrimination thresholds without inter-coefficient masking measured at our lab (Legge-like experimental data [16]). A frequency-dependent (one octave width) Gaussian kernel, fig. 2.c, has been heuristically introduced according to the results of refs. [9–12],

$$h_{ij} = k_i \, e^{-\frac{(f_i - f_j)^2}{\sigma(f_i)^2}} \tag{7}$$

where f_i is the spatial frequency meaning of the coefficient a_i, $\sigma(f_i)$ is the variable width of the kernel, $\sigma(f_i) = \frac{1}{3}|f_i| + 0.05$, with $|f|$ in cycl/deg, and k_i is a constant to obtain a unit-volume kernel.

For mathematical convenience (see section 5) a small linear term (proportional to $|a_i|$) has been included in the response model. This linear band-pass term (fig. 2.a) dominates for very low amplitudes. It is consistent with the fact that for low amplitude patterns the HVS response is roughly linear and it is well described by a band-pass function, the *Contrast Sensitivity Function* (CSF) [17].

In our implementation we have used a block DCT as a T transform, and no spatial interactions between neighbour blocks have been considered in h, but the analytical results presented can also include spatial interactions and can be applied to any wavelet-like transform.

3.3 Perceptual Metric using the Non-Linear Normalization

Gradient equation. Taking partial derivatives in the non-linear normalization model, eq. 6, we have the following gradient matrix:

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

The slope of the response has three contributions: two diagonal contributions and one off-diagonal contribution given by the interaction kernel. Note that from medium to high amplitudes the slope decreases with amplitude, i.e. the increase in the response is inhibited for high energy coefficients. Also note that the off-diagonal contribution is always negative, i.e. the increase in the response to one coefficient is also inhibited by high energy neighbour coefficients.

Perceptual Metric in the DCT domain. Figure 3.b shows the perceptual metric in the DCT domain that come from the expression 6 using the gradient of the non-linear normalization model, eq. 8.

The qualitative meaning of the metric elements (which give the relations between different coefficients of the feature vectors) depends on how the 2D DCTs are scanned to construct the 1D feature vectors. Figure 3.a explains the zigzag scanning that has been applied to the 2D DCTs. According with the zigzag scanning, the frequency meaning of the diagonal elements of h, ∇R, W, and Γ (when defined in the DCT domain) progresively increases from zero to the Nyquist frequency.

From fig. 3.b it is clear that the relative perceptual relevance of the DCT features highly depend on frequency (the diagonal of W has a low-pass shape), i.e. the DCT feature space is perceptually anisotropic. It is also clear that DCT features are not perceptually independent because W is not diagonal, i.e. the perceptually privileged directions of the DCT feature space are not aligned with

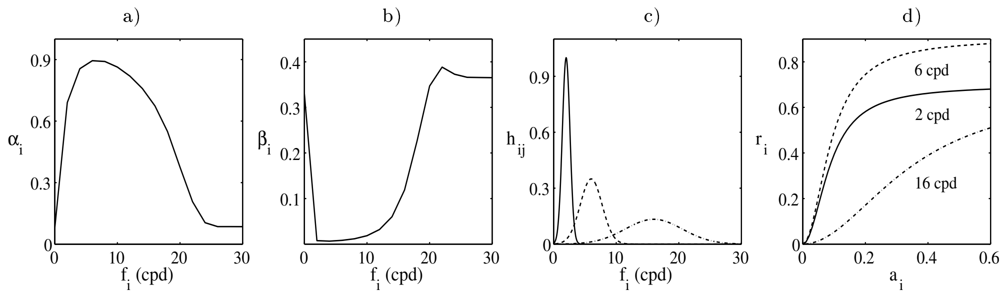

Fig. 2. Parameters of the vision model and non-linear response functions. The values in these figures assume that the amplitude of the coefficients is expressed in contrast (amplitude over mean luminance) which ranges between 0 and 1. The response examples of fig. 2.d show the basic (sigmoid) behaviour of eq. 6, but they are not general because the response to one coefficient depends on the background (depends on the value of the neighbour coefficients). These particular curves were computed for the particular case of no additional masking pattern (zero background).

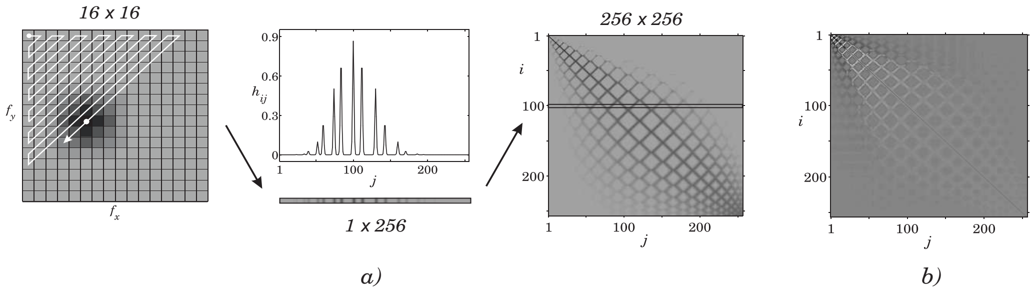

Fig. 3. DCT zigzag scanning and perceptual metric in the DCT domain. A convenient way to scan the 2D DCT is the *zigzag scheme* used in the JPEG standard [18] because it groups together coefficients with similar frequency. In the final vector the frequency progresively increases from the DC component in the first coefficient, up to the diagonal Nyquist frequency for the last coefficient –i.e. $(32, 32)$ *cpd* if the sampling frequency in each direction is 64 *cpd*–. In fig. 3.a the kernel for the coefficient $f_0 = (18, 12)$ *cpd* is zigzag scanned and introduced in its corresponding position in h. According to this scanning scheme, the coefficient of frequency f_0 is in the position $i = 100$, so the values that give the relations of a_{100} with a_j for $j = 1 \cdots 256$ form a raw vector that goes in the raw $i = 100$ to be applied on the input column vectors to give $(h{*}|a|^2)_{100}$. Fig. 3.b shows the perceptual metric for the average DCT. Dark and light pixels indicate positive and negative values respectively. An exponent has been applied to enhance the visibility of the off-diagonal elements.

the axis of the space. This implies that an additional transform is needed to remove the perceptual correlation between the DCT features.

Recall that the gradient (and then the metric) is input-dependent, so there are no global privileged directions in the space. This implies that the decorrelation transform must be local.

4 Joint Statistical and Perceptual Decorrelation through the Non-linear Normalization Model

In this work we propose the use of the *non-linear normalization model* as a feature decorrelation mapping from both statistical and perceptual points of view. First because it transforms the DCT domain in a perceptually Euclidean space, and second, because, its structure makes it a special form of predictive coder, therefore the output, r, should show less statistical correlation than the input DCT.

The basic idea of predictive coding (or DPCM) [18] is to remove from each coefficient the part that can be predicted from the neighbour coefficients. If a prediction of each coefficient is discounted from the original signal in some way, the cross-correlation between neighbour coefficients of the result will be highly reduced. In the commonly used DPCM the discount is linear: the prediction is substracted from the input giving a decorrelated error signal [18].

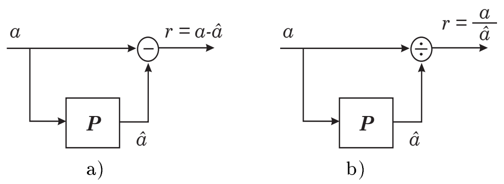

Fig. 4. Different DPCM schemes. Figure 4.a shows classical linear DPCM. Figure 4.b shows the DPCM interpretation of the non-linear normalization. In this case, the part which is predictable from the neighbour coefficients is discounted normalizing the input signal by the prediction.

It has been shown that the normalization of eq. 6 can be optimized to maximize the decorrelation of the output [9]. This normalization by a weighted sum of the neighbour coefficients can be interpreted as a (non-linear) divisive DPCM (see fig. 4): if the central point of the Gaussian kernel is set to zero (i.e. if the coefficient a_i is not taken into account in $(h * |a|^2)_i$ as is done in [9]), the convolution in the denominator can be seen as a prediction of the energy of each coefficient from the energies of its neighbours. The division will be a different way of discounting this prediction from the input.

In fact, the prediction stage in the *non-linear model II* is similar to the prediction scheme that has been succesfully used in [19] to exploit the conditional probabilities of the transform coefficients to encode them in a more efficient way. This suggest that the interaction scheme in *non-linear normalization* could certainly remove the statistical correlation in r. It is important to remark that the parameters of eq. 6 that will be used in this work are empirical, i.e. they haven't been specifically optimized to improve the decorrelation in a given trainning set.

5 Quasi-analytical Inversion of the Non-linear Normalization

Problem Statement. The interaction (and prediction) kernel which makes *non-linear normalization model* useful for statistical decorrelation also makes it non-invertible. 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 their convergence is not guaranteed and may be very sensitive to the initialization.

Quasi-analytical Inversion. In spite of the non-invertible nature of R, if a linear approximation is considered around a point a_a, the inverse function can be locally written as,

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

where the unknown gradient of the inverse function can be be related to the (known) gradient of the response: $\nabla R^{-1}(R(a_a)) = \nabla R(a_a)^{-1}$. 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_a, a_a), up to the desired point r_b (see fig. 5):

$$a_b = a_a + \int_{r_a}^{r_b} \nabla R^{-1}(r) \cdot dr \tag{10}$$

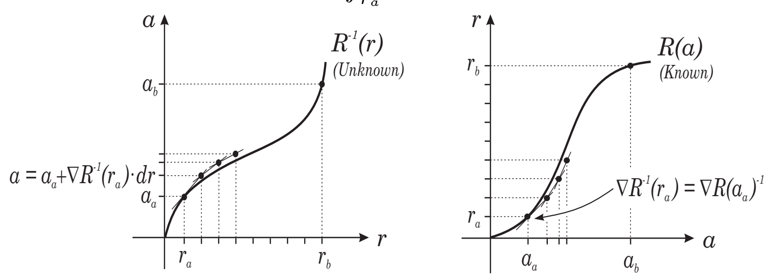

Fig. 5. Inverse computation integrating the increments of the inverse function. In each iteration, the unknown gradient is computed from the known response at that point.

The computation of the inverse is now analytically formulated in eq. 10. As this integral must be numerically solved we have called this method *quasi-analytical* in contrast to the numerical search-based methods.

Convergence of the Solution. The existence and uniqueness of the solution of an *initial value problem* is guaranteed if the gradient to be integrated, $\nabla R^{-1}(r)$, is bounded for every possible input [20]. In our case, $\nabla R(a)$ should not vanish anywhere. The linear term in the response gives a constant term in the gradient which ensures a non-zero slope for every a. This guarantees that the integration of eq. 10 is possible and always gives the appropriate solution.

To test the speed and robustness of the inverse computation, the DCT blocks of several natural images (Barbara, Einstein, Boats, Peppers, Golden Hill and Lena) were transformed according to the *non-linear normalization* and then inverted back using the eq. 10 solved with a 4^{th} order Runge-Kutta algorithm. The effect of the initialization and the number of integration steps was explored. Figure 6 shows the DCT reconstruction error as a function of the number of integration steps for two different blocks and different initial conditions.

The inversion experiments on several natural test images show the following trends:

- **The solution is always found**. The experiments confirm the theoretical existence and uniqueness result: the proposed method achieves the appropriate inverse (with negligible distortion), for every response block, no matter the initial conditions with a reasonably small number of integration steps (no more than 25 integration steps were used in any case).
- **Speed**. The majority of the responses ($\sim$90% in the explored images) appropriately converge to its corresponding DCT in 3-6 integration steps from

very different initial conditions. The solid line in fig. 6 is an example of this kind of blocks. However, we found that ~10% of the DCT blocks, usually corresponding to sharp spectrum regions, require a more accurate integration. The dashed line in fig. 6 represents the worst-behaved block of the trainning set.

– **Robustness**. The distortion in the inverse does not substantially depend on the initial conditions, but on the nature of the block (see fig. 6), so the algorithm is insensitive to the initialization. Generic $1/f$ or flat spectrums give good results for every block.

6 Decorrelation Experiments

The decorrelation properties of the proposed representation were compared with the standard PCA representation (i.e. the domain of eigenfunctions of Γ) and with the domain of eigenfunctions of W, which will be referred to as *Perceptual Principal Component Analysis* (PPCA). The local DCT which is the best fixed-basis approximation to PCA analysis for natural images [2] was also explored. The local DCT is also interesting because it is the first linear stage, T, in the proposed representation, (T, R), and the eventual improvements over the DCT will be given by the non-linear normalization R. The spatial representation has been included as a useful example of highly correlated domain.

The PCA representation was computed from the covariance around the average of a set of natural images (Barbara, Einstein, Boats, Peppers, Hill and Lena). The PPCA was computed from the average perceptual metric, originally defined over the DCT blocks of the trainning set and subsequently transformed to the other domains using eq. 6.

The second order relation matrices, Γ and W, in the different domains and the corresponding interaction measures, η_s and η_p, are shown in fig. 7 and table 1.

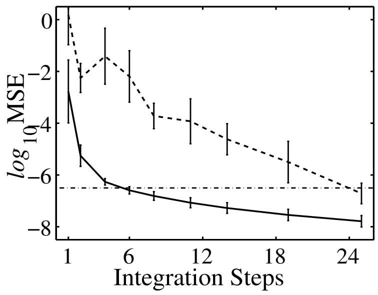

Fig. 6. Reconstruction errors of a *tipical* block (solid line) and a *difficult* block (dashed line). The curves shown are the average over several initial conditions: the mean DCT over the trainning set, a $1/f$ spectrum DCT and a flat spectrum DCT. The deviations bars indicate the dispersion in the final distortion due to the different initial conditions. The differences below the straigth (dashdot) line are visually negligible.

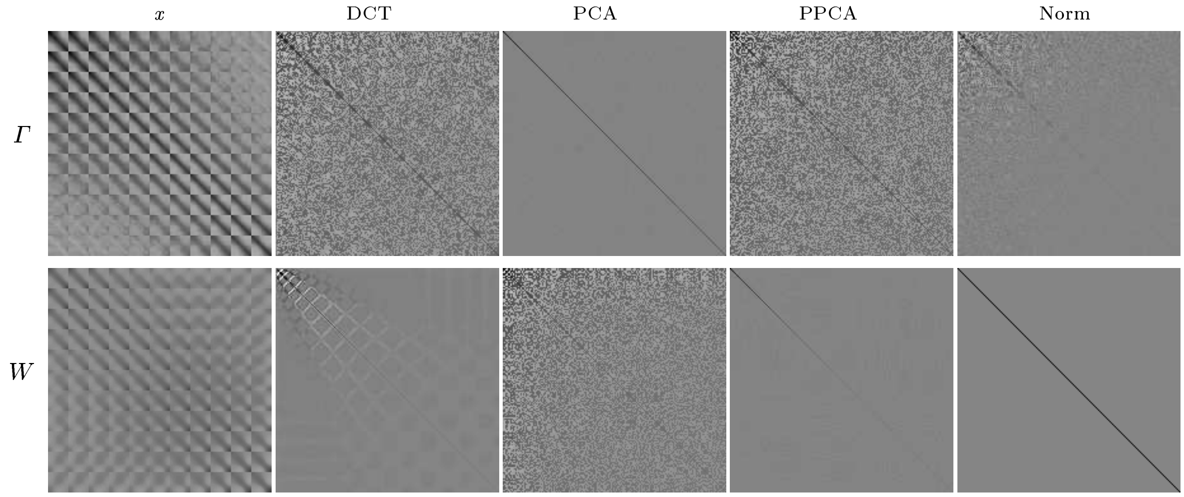

Fig. 7. Covariance (upper row) and perceptual metric (lower row) in different domains. For the sake of clarity only the upper-left 176×176 submatrix is shown. The matrices in the spatial domain are the result of a raster scanning. The matrices in the other transform domains are the result of a zig-zag scanning of the DCT domain. The frequency values of the displayed elements in the DCT domain range from 0 to 26 *cpd*.

The higly non-diagonal nature of W in the spatial domain shown in these results represent an additional argument angaist the spatial domain representation that complements the classical reasonings exclusively based on the non-diagonal nature of the covariance matrix [1, 2].

The DCT domain certainly reduces the statistical and perceptual interactions by an order of magnitude with regard to the spatial domain but it still doesn't completely remove none of them.

The linear approaches that only take into account one of the (statistical or perceptual) relations, PCA or PPCA, are not acceptable because, in this case, the other relation is increased in the resulting representation.

The proposed representation, DCT plus non-linear normalization transform, gives the best results. On one hand, it achieves a complete perceptual decorrelation for every input. It must be stressed that it works with local (not average) W, so, in this sense the perceptual decorrelation is better than in the PPCA or any other PCA-like approach. On the other hand, the statistical interaction is also highly reduced, almost an order of magnitude with regard to the DCT.

Table 1. Feature interaction measures in different representation domains

	x	DCT	PCA	PPCA	Norm.
η_s	41.5	3.9	0.0	4.4	0.5
η_p	59.8	1.5	6.0	0.0	0.0

Of course, Γ and W can be combined to design a linear transform which simultaneously gives $\eta_s = \eta_p = 0$ [13]. This alternative linear input-dependent transform has not been included in this discussion because it has a practical disadventage (as well as PCA and PPCA): if the ad-hoc basis functions are locally adapted, they need to be storaged to invert the representation. On the contrary, the proposed representation is adaptive, but the basis does not need to be storaged because the representation is invertible.

7 Concluding Remarks

In this paper the perceptual correlation between the features of an image representation has been formalized through the perceptual metric matrix in the same way as the statistical correlation is described by the covariance matrix.

We have presented a perceptually inspired image representation that simultaneously reduces the statistical and perceptual correlation between the features. It first uses a linear local frequency transform and after a non-linear energy normalization is applied to the local DCT coefficients. The good statistical behaviour of this non-linear perceptual model relies on its divisive-DPCM structure. The proposed representation improves the decorrelation properties of a fixed basis representation such as the DCT without the basis storage problem of linear input-dependent PCA-like transforms because an efficient method to invert it has been presented.

According to the results presented here, the non-linear mapping R may be a very interesting second stage after the linear DCT or wavelet-like transforms used in many image analysis applications. Recent results in image transform coding [21] confirm this point.

References

1. K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, Boston, 1990.
2. R.J. Clarke. *Transform Coding of Images.* Academic Press, NY, 1985.
3. A. Bimbo. *Visual Information Retrieval.* Morgan Kaufman Publishers, NY, 1999.
4. H. Li, B.S. Manjunath, and S.K. Mitra. Multisensor image fusion using the wavelet trasnformation. *Graph. Mod. Im. Proc.*, 57(3):235–245, 1995.
5. D. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394, 1987.
6. J.G. Daugman. Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Tran. Biomed. Eng.*, 36:107–114, 1989.
7. B.A. Olshausen and D. Field. Emergence of simple-cell receptive field properties by a learning a sparse code for natural images. *Nature*, 381:607–609, 1996.
8. A.J. Bell and T.J. Sejnowski. The independent components of natural scenes are edge filters. *Vision Res.*, 37(23):3327–3338, 1997.
9. E.P. Simoncelli and O. Schwartz. Modeling surround suppression in V1 neurons with a statistically-derived normalization model. In M.S. Kearns, editor, *Adv. in Neural Inf. Proc. Syst.*, volume 11. MIT Press, 1999.

10. J.A. Solomon, A.B. Watson, and A. Ahumada. Visibility of DCT basis functions: Effects of contrast masking. *IEEE Proc. Data Compress. Conf.*, pages 361–370, 1994.
11. P.C. Teo and D.J. Heeger. Perceptual image distortion. *Proc. of the First IEEE Intl. Conf. Im. Proc.*, 2:982–986, 1994.
12. A.B. Watson and J.A. Solomon. A model of visual contrast gain control and pattern masking. *J. Opt. Soc. Am. A*, 14:2379–2391, 1997.
13. I. Epifanio and J. Malo. Linear transform for simultaneous diaginalization of covariance and perceptual metric in image coding. Technical Report 3, Grup de Visió. Dpt. d'Informàtica, Universitat de València, 2000.
14. B. Dubrovin, S. Novikov, and A. Fomenko. *Modern Geometry*. Springer Verlag, NY, 1982.
15. A.M. Pons, J. Malo, J.M. Artigas, and P. Capilla. Image quality metric based on multidimensional contrast perception models. *Displays*, 20:93–110, 1999.
16. G.E Legge and J.M. Foley. Contrast masking in human vision. *J. Opt. Soc. Am.*, 70:1458–1471, 1980.
17. B.A. Wandell. *Foundations of Vision*. Sinauer Assoc. Publish., Massachusetts, 1995.
18. A.M. Tekalp. *Digital Video Processing*. Prentice Hall, Upper Saddle River, NJ, 1995.
19. R.W. Buccigrossi and E.P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans. Im. Proc.*, 8(12):1688–1701, 1999.
20. D.J. Logan. *Introduction to non-linear partial differential equations*. Wiley&Sons, NY, 1994.
21. J. Malo, F. Ferri, R. Navarro, and R. Valerio. Perceptually and statistically decorrelated features for image representation: Application to transform coding. *Proc. XV Intl. Conf. Patt. Recog.*, 2000.