# Perceptually uniform quantization versus average error minimization in image transform coding

Electronics Letters, Vol. 35, 13, 1067–1068 (1999)

J. Malo*, F.J. Ferri†, J. Albert† and J. Soret†

(*) Dept. d'Òptica
(†) Dept. d'Informàtica i Electrònica
Universitat de València
Dr. Moliner, 50. 46100 Burjassot, València. Spain
jesus.malo@uv.es, ferri@uv.es

An alternative transform coder design criterion based on restricting the *maximum perceptual error* of each coefficient is proposed. This perceptually uniform quantization of the transform domain ensures that the perceptual error will be below a certain limit regardless of the particular input image. The results show that the proposed criterion improves the subjective quality of the conventional *average error* criterion even if it is weighted with the same perceptual metric.

This work has been supported by CICYT projects 1FD97-0279 and TIC 98-677-C02-02.

## Introduction.

In the transform coding paradigm, the input image, $\mathbf{A}$, is mapped onto a local frequency space, $\mathbf{a} = T(\mathbf{A})$, with $\mathbf{a} = \{a_f\}_{f=1}^{m}$. The signal is then simplified through the quantization of each transform coefficient, $a_f$. Two main problems have to be solved once a certain transform is given. First, one must decide how to distribute quantization levels in the amplitude range of each $a_f$ (1D *quantizer design*). And second, one must decide how many quantization levels are assigned to each $a_f$ (*bit allocation*). Conventionally, the minimization of the *average* quantization error over a training set [1] has been used to deal with these problems. This error must contain some perceptually meaningful metric to take into account the human visual system (HVS) characteristics [2, 3]. However, even with this perceptually corrected error, the conventional techniques may arbitrarily underperform on some individual images [2]. In fact, obtaining a minimum *average* quantization error lead to the accumulation of quantization levels in particular regions. This may give rise to important subjective deficiencies on patterns which are not sufficiently represented in the training set.

In this work, a fully-perceptual design criterion, the *maximum perceptual error* criterion (MPE), is presented as an alternative to the (eventually perceptual) *mean square error* criterion (MSE).

## Frequency and amplitude-dependent perceptual metric.

In image coding applications, an input-independent metric based on a linear band-pass filter (the *Contrast Sensitivity Function*, CSF) is usually applied [2, 3]. However, the HVS response is highly non-linear and the CSF-based model must be extended to include the amplitude non-linearities in order to obtain adaptive metrics. Here, a recently proposed metric [4] will be assumed,

$$W_f(a_f) = \left( CSF_f^{\,-1} + a_f\, G_f(a_f) \right)^{-2} \tag{1}$$

where $G_f(a_f)$ are empirical monotonically increasing functions of amplitude for each spatial frequency that account for the HVS non-linearities and the contrast masking effect [4]. This metric simplifies to the square of the CSF for very low amplitudes, so CSF-based metrics are obtained as a particular (linear) case valid in certain conditions. Therefore, the perceptual difference between two close patterns is given by:

$$D(\mathbf{a}, \mathbf{a} + \Delta\mathbf{a})^2 \;=\; \sum_{f=1}^{m} D_f^2 \;=\; \sum_{f=1}^{m} W_f(a_f)\, \Delta a_f^2 \tag{2}$$

## Perceptually-weighted Mean Square Error results.

Using the perceptual metric of eq. 1 and the asymptotic results for general non-euclidean metrics [5], the optimal (MSE) point density can be obtained by:

$$\lambda_{f\,opt}(a_f) \propto (W_f(a_f)\, p(a_f))^{1/3} \tag{3}$$

This result only differs in the factor $W_f(a_f)$ from the well-known euclidean result, which is proportional to the cube root of the probability density function, $p(a_f)$ [1]. Once the quantizer design problem is solved, the number of levels, $N_f$, can be obtained from the optimal *average perceptual error* for each coefficient [1]. In our case,

$$\overline{D}^2_{f\,opt} = \frac{\sigma_f^2}{12N_f^2}\left(\int (W_f(\sigma_f a_f)\tilde{p}(a_f))^{1/3} da_f\right)^3 = \frac{\sigma_f^2}{N_f^2} H_f \tag{4}$$

where $\tilde{p}(a_f)$ is the normalized unit-variance *pdf* of the coefficient. The variance, $\sigma_f^2$, and the metric-dependent parameter, $H_f$, can be seen as the weighting factors that control the bit allocation for each coefficient.

## The Maximum Perceptual Error criterion.

In order to restrict the MPE throughout the amplitude range of each coefficient, a perceptually uniform distribution of quantization levels must be achieved. In this way, individual distortions will be bounded (by a constant value) regardless of the statistical population of certain amplitude values in certain coefficients.
If the coefficient $f$ is quantized through $N_f$ levels distributed according to a density, $\lambda_f(a_f)$, the maximum (euclidean) error at an amplitude, $a_f$, will be bounded by the step size, which is inversely proportional to $N_f \lambda_f(a_f)$. Therefore, the *maximum perceptual error* at that amplitude is $\widehat{D}_f^2 = W_f(a_f) \cdot (2N_f\, \lambda_f(a_f))^{-2}$. To make this bound constant over the amplitude range, the optimal (MPE) density should follow:

$$\lambda_{f\,opt}(a_f) \propto W_f(a_f)^{1/2} \tag{5}$$

and $N_f$ will be obtained as in eq. 4 from the optimal *maximum perceptual error*:

$$\widehat{D}^2_{f\,opt} = \frac{1}{4N_f^2}\left(\int W_f(a_f)^{1/2} da_f\right)^2 = \frac{1}{N_f^2} H'_f \tag{6}$$

If a simple linear perception model is assumed (neglecting the amplitude-dependent term in eq. 1), the proposed formulation leads to uniform quantizers and $N_f \propto CSF_f$, just as in the JPEG standard [6]. However, when considering the non-linear behaviour, the quantization step size becomes non-uniform and the bit allocation is changed. Similar results can be obtained through an intuitive approach based on perceptual data of contrast just noticeable differences [7]. Eqs 3, 4 and 5, 6 constitute the key facts that control the coder design for MSE and MPE criteria, respectively. It is worth noting that the main difference stems from the fact that there is no dependence on image statistics in eqs. 5 and 6.

## Results and Discussion.

Several experiments were carried out using the above approaches in the DCT domain, with and without the non-linear term in eq. 1. In the MSE cases, the $16 \times 16$ low-rate quantizers were implemented from the asymptotic results through the LBG method [1], using a training set of 15360 blocks from 8-bit head-and-shoulder images. In the non-linear MPE case, the 1D quantizers were obtained through the corresponding companding functions: the cumulative $\lambda_f$. The bit allocation in all schemes was obtained from eqs. 4 or 6 through a greedy integer-constrained algorithm [1].

The performance of the MSE approach (even when taking into account amplitude non-linearities) is very similar to the linear MPE (JPEG-like) approach. In contrast, the non-linear MPE approach clearly gives the best subjective results at the same compression ratios, especially in the high frequency details. Fig. **1** shows an example of the results for three different quantizer design criteria (at 0.5 bpp). Table **1** show some distortion measures for the images of fig. **1**. Fig. **2** illustrates the way in which MSE and MPE criteria allocate the available quantization levels using the same non-linear metric. The MSE encoding effort is focused in the low amplitude regions (because image statistics) leading to significant differences with regard to MPE (especially for medium and high frequencies). While differences between the two MPE approaches are due to the different accuracy of the models used [7], the (also significant) differences with regard to the non-linear MSE are due only to the use of different criteria. The results suggest that the MSE criterion cannot fully exploit the eventual adventages of sophisticated non-linear metrics. Despite the non-linear MPE images do not minimize the perceptual MSE (see the example on table **1**) they have the best subjective quality. This behaviour indicates that ensuring a bounded perceptual distortion in the particular image to be encoded may be more important than minimizing the *average* perceptual error over a set of images.

## References


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

[2] B. Macq. Weighted optimum bit allocations to orthogonal transforms for picture coding. *IEEE SAC*, 10(5):875–883, 1992.

[3] N.B. Nill. A visual model weighted cosine transform for image compression and quality assessment. *IEEE Trans. Comm.*, 33:551–557, 1985.

[4] J. Malo, A.M. Pons, and J.M. Artigas. Subjective image fidelity metric based on bit allocation of the human visual system in the DCT domain. *Im. Vis. Comp.*, 15:535–548, 1997.

[5] Y. Yamada, S. Tazaki, and R.M. Gray. Asymptotic performance of block quantizers with difference distortion measures. *IEEE Trans. Inf. Theor.*, 26(1):6–14, 1980.

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

[7] J. Malo, A.M. Pons, and J.M. Artigas. Bit allocation algorithm for codebook design in vector quantization fully based on human visual system non-linearities for suprathreshold contrasts. *Electr. Lett.*, 31:1229–1231, 1995.


**Table 1**: Distortion measures of the different images

| | Distortions | | |
|---|---|---|---|
| Images | SNR | MSE | MPE |
| Non-linear MSE | 4.19 | 646.8 | 151.7 |
| Linear MPE (JPEG) | 4.08 | 681.1 | 185.9 |
| Non-linear MPE | 4.92 | 664.0 | 118.4 |

**Fig. 1**: *From top-left to bottom-right: original, non-linear MSE, linear MPE (CSF-based JPEG), and non-linear MPE.*

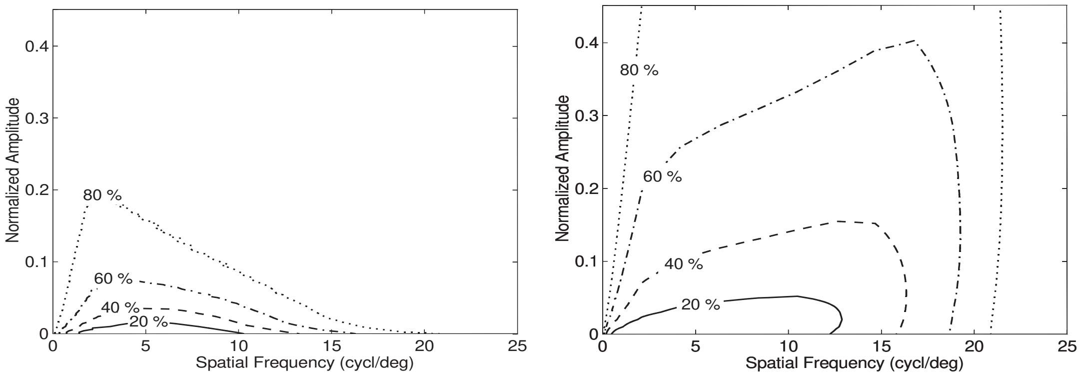


**Fig. 2**: *Distribution of quantization levels (in % of the total number) in the frequency-amplitude plane using MSE (left) and MPE (right) with the same nonlinear metric.*