The role of perceptual contrast non-linearities in image transform quantization [1]

J. Malo [a], F. Ferri [b], J. Albert [b], J. Soret [b] and J.M. Artigas [a]

[a] *Dpt. d'Òptica*

[b] *Dpt. d'Informàtica i Electrònica*

Universitat de València

C/ Dr. Moliner 50. 46100. Burjassot, València, Spain

e-mail: Jesus.Malo@uv.es

The conventional quantizer design based on *average error* minimization over a training set does not guarantee a good subjective behaviour on individual images even if perceptual metrics are used.
In this work a novel criterion for transform coder design is analyzed in depth. Its aim is to bound the perceptual distortion in each individual quantization according to a non-linear model of early human vision. A common comparison framework is presented to describe the qualitative behaviour of the optimal quantizers under the proposed criterion and the conventional rate-distortion based criterion. Several underlying metrics, with and without perceptual non-linearities, are used with both criteria. Analytical results show that the proposed design criterion gives rise to a JPEG-like quantization if a simple linear metric is used. Experimental results show that significant improvements over the perceptually weighted rate-distortion approach are obtained if a more meaningful non-linear metric is used.

Key words: Image Coding. Perceptual Quantization. Non-linear Perception Model.

1 Introduction

According to the well-known image transform coding paradigm, an appropriate transformation of the signal is carried out prior to the quantization stage to simplify the quantizer design. Given a certain transform (usually a local frequency transform), two intimately related problems must be solved. The first problem is to decide how to distribute quantization levels in the amplitude

[1] Partially supported by CICYT projects TIC98-677-C02-02 and TIC 1FD97-279.

range of each transform coefficient. The second problem is to determine the number of quantization levels that will be devoted to each coefficient. The first is usually known as *1D quantizer design* and the second is often referred to as *bit allocation*.

It is widely accepted that for image coding applications that are judged by a human observer, the properties and limitations of the human visual system (HVS) have to be taken into account in order to achieve better subjective results [1]. In transform-based coding, the sensitivity of the human viewer to the transform basis functions should be taken into account to adapt the quantization step size for each coefficient to minimize the perceived distortion [2, 3]. As stated by Watson [**?**], there are two ways to exploit the HVS characteristics in image coding: *a)* the usual *signal-based* way, which consists of adapting the conventional approach based on minimizing the *average error* [4] through the appropriate perceptual distortion metric, and *b)* an alternative *fully-perceptual* way, which mimics the way the HVS encodes the visual information.

The HVS characteristics been applied to different approaches in a number of ways. While the HVS sensitivity frequency dependency has been widely used for bit allocation [4–14], the corresponding amplitude dependencies have rarely been used, and only under the fully-perceptual paradigm [15–17].

Signal-based quantizers that minimize the mean square error have some drawbacks for individual images [18] even if the error measure is perceptually weighted [10]. This is mainly because the statistics of the training set may bias the encoding effort towards the most populated regions in order to minimize the average distortion. Under these conditions some values of amplitude for certain coefficients of importance for the HVS may become badly represented. This effect is particularly important at high compression ratios when the number of reproduction levels is dramatically reduced and the probability of badly represented amplitudes increases. This suggests that the use of an *average error* criterion may not be the best choice given a particular HVS model.

The objective of this work is to assess the relative merits of a perceptually weighted signal-based coding scheme when a HVS model including frequency and amplitude dependencies is considered. To show that this is not the best *subjective* way of solving the problem, an alternative to the weighted MSE criterion is proposed: the *maximum perceptual error* (MPE) criterion. While the MSE implies a minimization over a training set, the MPE consists of setting limits on the amount of distortion permitted for every frequency and amplitude taking into account the HVS.

The main goal is to compare these two design criteria when using different HVS models. Moreover, the particular effects of introducing the amplitude non-linearities is analyzed in depth in both cases.

The paper is organized as follows. The selected perceptual metric including frequency and amplitude properties of the HVS is introduced in Section 2. In section 3, this metric is used to obtain particular quantizers from the asymptotic results of the rate-distortion theory, based on the MSE criterion. The alternative MPE criterion is presented in section 4. In section 5, the performance of the optimal quantizers using both design criteria with the same perceptual metrics is compared. To conclude, some final remarks are presented in section 6.

2 Frequency and amplitude-dependent perceptual metric

According to currently accepted perception models [19–22], the HVS maps the input spatial patterns, $\mathbf{A}$, onto a frequency feature space through a set of m local frequency filters, $\mathbf{a} = T[\mathbf{A}]$, with components a_f, where $f = 1, \ldots, m$. Then, a non-linear log-like transform, R, is applied to yield the so-called response representation, $\mathbf{r} = R[\mathbf{a}]$, with components r_f.

The perceptual metric in the feature space, $W(\mathbf{a})$, can be obtained from an appropriate contrast response model. The metric is related to the non-linearities of the response R and, under certain assumptions, can be directly computed from its gradient as $W(\mathbf{a}) = \nabla R(\mathbf{a})^T \cdot \nabla R(\mathbf{a})$ [23]. It can also be empirically obtained from the HVS discrimination abilities. In the feature space, two patterns are perceived as different if the distance between them is above a certain discrimination threshold value, τ. The vector difference between two just distinguishable patterns, $\Delta\mathbf{a}^\star(\mathbf{a})$, is called *just noticeable difference* (JND), and the square perceptual distance, D^2, under the *ideal observer assumption* [21, 22] can be written as:

$$D(\mathbf{a}, \mathbf{a} + \Delta\mathbf{a}^\star(\mathbf{a}))^2 \;=\; \Delta\mathbf{a}^\star(\mathbf{a})^T \cdot W(\mathbf{a}) \cdot \Delta\mathbf{a}^\star(\mathbf{a}) \;=\; \tau^2 \tag{1}$$

All the elements of the metric can be obtained with an appropriate number of JND data at a point, $\mathbf{a}$. If the interaction among the filter outputs is neglected, W will be diagonal. In this case, each non-zero element, W_f, will not depend on $a_{f'}$ for $f' \neq f$ [21, 23]. From eq. 1, each diagonal element will be inversely proportional to the square of the JND in the corresponding axis. Using a fit of the amplitude JND data, a previously tested metric [23, 24] is obtained:

$$W_f(a_f) = \tau^2 \Delta a_f^\star(a_f)^{-2} = \tau^2 \left(CSF_f{}^{-1} + \frac{\frac{a_f}{L}\left(k_f\left(\frac{a_f}{L}\right)^{n_f} - CSF_f{}^{-1}\right)}{(k_f CSF_f)^{-1/n_f} + \frac{a_f}{L}} \right)^{-2} \tag{2}$$

where L is the local mean luminance, CSF_f is the *Contrast Sensitivity Function*, which can be computed with the expressions given by Kelly [25], Nill [26] or Nygan [12, 13], and k_f and n_f are the empirical dependency on frequency of the JND parameters [24].

It is worth pointing out that eq. 2 includes the effect of non-linearities as a correction of the linear threshold behavior represented by the CSF band-pass filter. If the threshold behavior, $a_f \to 0$, is assumed to be valid for suprathreshold amplitudes (as is done in simple linear models) the amplitude dependent term in eq. 2 vanishes and the CSF-based metric [26, 27] is obtained.

If a given coding scheme uses some particular transform basis, specific incremental threshold data should be used [2, 3]. However, as the impulse responses of the perceptual filter bank, T, closely resemble the basis functions used in image coding transforms [28, 29] a similar behavior can be expected [2]. Therefore the metric W can be considered a reasonable approximation with any generic local frequency transform. In some cases, it is also possible to obtain the behavior in the desired domain from the functions defined in other domains [26, 31].

3 Coder design through perceptual MSE minimization

The standard approach to transform coder design is rooted at the minimization of the mean square error, $\overline{D}_f^2$, between the original amplitude of each transform coefficient, a_f, and its quantized version. Using a generic perceptual metric, W, to measure the elemental distortions, it holds that:

$$\overline{D}_f^2 = \sum_{i=1}^{N_f} \int_{R_i} (a_f - {a_f}^i)^2 \, W_f({a_f}^i) \, p(a_f) \, da_f \tag{3}$$

where N_f is the number of quantization levels for that coefficient, ${a_f}^i$ is the i-th quantization level, R_i is the quantization region corresponding to the i-th level and $p(a_f)$ is the probability density function (*pdf*) of a_f.

Using the asymptotic results of Gish et al. [32] and Yamada et al. [33] for general non-euclidean metrics, the Bennett distortion integral [34] can be obtained in this case:

$$\overline{D}_f^2 = \frac{1}{12 N_f^2} \int \frac{W_f(a_f) \, p(a_f)}{\lambda_f(a_f)^2} \, da_f \tag{4}$$

[2] For instance, the same kind of amplitude non-linearity or absolute detection threshold for sinusoids and DCT basis functions is usually assumed [2, 5, 16, 30].

This expression relates the average distortion of a coefficient using the proposed metric with the density of quantization levels for that coefficient, $\lambda_f(a_f)$. The optimal 1D quantizer with the proposed metric can be obtained using the Hölder inequality in the standard way [4]:

$$\lambda_{f\,opt}(a_f) = \frac{(W_f(a_f)\,p(a_f))^{1/3}}{\int (W_f(a_f)\,p(a_f))^{1/3}\,da_f} \tag{5}$$

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 *pdf* [4, 35, 36].

The inclusion of the amplitude dependence of the metric also has significant effects on the bit allocation results. As the global squared distortion is given by the sum of the squared distortions coming from each coefficient, the individual distortions should be kept below a certain threshold in order to keep the global distortion as low as possible. This requirement determines a non-uniform bit allocation in the transform domain because more bits should be used to encode the more demanding coefficients [4, 37]. The intrinsic demand of each coefficient is given by its contribution to the total distortion. Substituting $\lambda_{f\,opt}$ in equation 4, the distortion for each optimal quantizer is:

$$\overline{D}^2_{f\,opt} = \frac{\sigma_f^2}{12N_f^2}\left(\int (W_f(\sigma_f a_f)\,\tilde{p}(a_f))^{1/3}\,da_f\right)^3 = \frac{\sigma_f^2}{N_f^2}\cdot H_f \tag{6}$$

where $\tilde{p}(a_f)$ is the normalized unit-variance *pdf* of the coefficient; the variance, σ_f^2, and the metric dependent parameter, H_f, can be seen as the weighting factors that control the intrinsic relevance of each coefficient. The optimal bit allocation is obtained imposing the same average distortion for each f [4]. In our case, from the optimal distortion results of equation 6, we obtain:

$$b_{f\,opt} = \frac{B}{m} + \frac{1}{2}\log_2(\sigma_f^2\cdot H_f) - \frac{1}{2m}\sum_{f=1}^{m}\log_2(\sigma_f^2\cdot H_f) \tag{7}$$

where $b_f = \log_2 N_f$ is the number of bits used to encode the coefficient f; and B is the total number of available bits.

The final effect of the amplitude non-linearities is to weight the *pdf* in the 1D quantizer design and to introduce a new factor in the bit allocation expressions. These expressions generalize the previously reported results in the context of image transform coding [4, 9, 10] to include amplitude dependencies of the perceptual metric.

If only the frequency dependency of HVS sensitivity is taken into account (as is done in simple linear models), the metric $W_f(a_f)$ will just be a function of frequency related to the CSF. Thus, W_f can be taken from the integrals in eqs. 5 and 6 to give an explicit perceptual frequency weight term in the bit allocation expression [4, 9, 10]. In this case, the 1D quantizers will only depend on the signal statistics, and the density of quantization levels is then proportional to the cube root of the *pdf* [4, 35, 36]. Band-pass weights have been heuristically used in several coding schemes [6, 11, 12] to introduce simple HVS features in the euclidean (purely statistical) design.

Although all these results only hold in the high resolution case, these analytical expressions clarify the effect of the the metric and often turn out to be a reasonable approximation even in the medium to low resolution cases.

4 Coder design through Maximum Perceptual Error bounding

In the different MSE-based approaches, the final quantizer may show a perceptually uneven distribution of the quantization levels depending on the statistics of the training set. The accumulation of quantization levels in the more probable regions which is performed to minimize the average perceptual error does not ensure good behavior on a particular image.

In order to prevent high perceptual errors on individual images coming from outlier coefficient values, the coder should be designed to bound the maximum perceptual error (MPE) for every frequency and amplitude. While the MSE criterion minimizes the *average* distortion, $\overline{D}_f^2$, the proposed MPE criterion bounds the *maximum individual* distortion, $\widehat{D}_f^2(a_f)$.

If a given coefficient is represented by N_f quantization levels distributed according to a density, $\lambda_f(a_f)$, the maximum euclidean quantization error at an amplitude, a_f, will be bounded by half the euclidean distance between two levels:

$$\Delta a_f(a_f) \leq \frac{1}{2N_f\,\lambda_f(a_f)} \tag{8}$$

Assuming a generic diagonal frequency and amplitude-dependent metric, the MPE at that amplitude will be related to the metric and the density of levels:

$$\widehat{D}_f^2(a_f) = \frac{1}{2N_f\,\lambda_f(a_f)} \cdot W_f(a_f) \cdot \frac{1}{2N_f\,\lambda_f(a_f)} = \frac{W_f(a_f)}{4N_f^2\,\lambda_f^2(a_f)} \tag{9}$$

According to this, the maximum perceptual distortion bound is constant over the amplitude range only if the point density varies as the square root of the metric,

$$\lambda_{f\,opt}(a_f) \;=\; \frac{W_f(a_f)^{1/2}}{\int W_f(a_f)^{1/2} da_f} \tag{10}$$

With these optimal densities, the maximum perceptual distortion in each coefficient will depend on the number of allocated levels and on the integrated value of the metric:

$$\widehat{D}^2_{f\,opt} \;=\; \frac{1}{4N_f^2}\left(\int W_f(a_f)^{1/2} da_f\right)^2 \tag{11}$$

Fixing the same maximum distortion for each coefficient, $\widehat{D}^2_{f\,opt} = k^2$, and solving for N_f, the optimal bit allocation is obtained:

$$b_{f\,opt} = \log_2 N_f = \log_2\left(\frac{1}{2k}\int W_f(a_f)^{1/2}\, da_f\right) \tag{12}$$

The expressions 10-12 constitute an alternative to the expressions 5-7 that are optimal in an average sense. The MPE criterion implies a perceptually uniform distribution of the available quantization levels in the feature space, and this only depends on the perceptual metric and not on the signal statistics.

A practical side effect of the fully-perceptual MPE criterion is that it does not need an off-line training stage. By its nature, it should be well-behaved over a wide range of natural images. The independence of the statistics of the images to be encoded should not be considered as rigidity: if the underlying perception model is adaptive, as are non-linear models with regard to linear models, the final quantization will be input-dependent.

It is interesting to point out that the optimal 1D quantizers in the MSE sense, eq. 5, reduce to the MPE optimal quantizer of eq. 10, if $p(a_f) \propto W_f(a_f)^{1/2}$. The differences in the placement of quantization levels will depend on the shape of the *pdf*s with regard to the shape of the metric. *Pdf*s which are too sharp may bias the distribution of quantization levels giving rise to a poor representation of some potentially important regions from a perceptual point of view. As a consequence, some subjectively relevant details of a particular image may be lost with the MSE approach even if the appropriate perceptual metric is taken into account.

The presented MPE formulation is qualitatively related to other fully perceptual approaches that have also taken the perceptual amplitude non-linearities of the HVS into account [15–17, 38]. In particular, the qualitative quantizer design of refs. [17, 38] is ruled by a sensitivity function, the *Information Allocation Function* (IAF), which is inversely proportional to the JNDs as the square root of the perceptual metric of section 2.

The equivalence between these fully perceptual approaches can be seen in their common background perception model, in the similarity of their underlying functions [3], and in their similar applications: these models have been applied to define perceptually meaningful distortion metrics [23, 24, 30, 40]. The IAF model has also been successfully used to perceptually match other signal processing algorithms, such as adaptive motion estimation [38, 41].

A particularly interesting quantizer is obtained after neglecting the amplitude-dependent term in the metric (eq. 2). In this case, uniform quantizers are obtained for each coefficient. The number of quantization levels per coefficient is simply proportional to the CSF, which is one of the recommended options included in the JPEG standard [5, 7, 8, 14].

5 Results and discussion

Experiments to compare the results of the different schemes for natural images at different compression ratios were carried out. A total of five quantizers were compared to explore the effect of perceptual amplitude non-linearities under the different design criteria:

- MSE-E: mean square error criterion with a euclidean metric, the standard *Mean Square Error* approach [4].
- MSE-F: mean square error criterion with a frequency-dependent perceptual metric [4, 6, 9–12].
- MSE-FA: mean square error criterion with a frequency and amplitude-dependent perceptual metric.
- MPE-F: maximum perceptual error criterion with a frequency-dependent perceptual metric. (the CSF-based JPEG-like scheme) [5, 7, 8, 14]
- MPE-FA: maximum perceptual error criterion with a frequency and amplitude-dependent perceptual metric.

In the MSE approaches, the perceptual features are gradually introduced to the average squared distortion measure combined with the statistical properties of

[3] Watson [16] uses the Legge fit of the incremental threshold data of gratings [39]. The noise-adaptive CSF used in ref. [15] is obtained from incremental thresholds of gratings over white noise and is quite similar to the IAF obtained from contrast incremental thresholds of gratings (see Fig.4.a of [15] and Fig.1 of [24]).

the signal. In the MPE approaches, the difference consists of whether or not the amplitude non-linearities are considered.

5.1 Implementation details

The training set included 60, 256×256, 8-bit monochrome video conference-like images (head-and-shoulders). The quantization block size was set to 16×16, so 15360 sample blocks were considered to estimate the amplitude *pdfs* of the MSE approach. Figure 1 shows the relevant statistical factors in the MSE approach. Figure 1.a shows the pdfs for some low, medium and high frequency coefficients (4, 8 and 16 cycl/deg). Figure 1.b shows the variance of each coefficient, σ_f^2. The factor H_f in the MSE-E case, eq. 6, depends only on the shape of the normalized unit-variance pdfs and is almost frequency-independent as reported in the literature [4].

The analytical results of sections 3 and 4 are strictly applicable under the high rate approximation [4, 34] to ensure that the pdfs and the metric are smooth enough to be considered constant in each quantization step. To design the actual low resolution MSE quantizers for the experiments, the LBG method [35, 42] was used with the appropriate metric. In each case, the iterative method was initialized with the asymptotic results. The final quantizers were quite consistent with the asymptotic assumptions. In the MPE-FA case, the 1D quantizer for each frequency was obtained from N_f uniformly distributed quantization levels mapped to the amplitude domain through the inverse of the proper companding function [4, 17, 38]. In the JPEG-like case, the quantization step size was set constant.

The actual bit allocation in all schemes was obtained through a greedy integer-constrained algorithm [4] based on the sequential allocation of one level to the coefficient with the largest distortion in each iteration. In the MSE and MPE cases, the distortion is given by the eqs. 6 and 11, respectively. The DC coefficient is always separately encoded using DPCM and will not be considered throughout the discussion sections.

5.2 Quantizer results

Figure 2 shows the bit allocation results for the compared approaches. In the MSE-E approach, the bit allocation is directly given by the variance. In the MSE-F approach, the CSF-based frequency weight cuts the high frequency tail of the variance curve. The MSE-FA reduces this restriction to a certain extent because the consideration of the amplitude non-linearities widens the band of perceptual interest. The same effect can also be seen in the fully perceptual approaches: while the MPE-F bit allocation is more concentrated in the low frequency region, the MPE-FA does include the high frequency region. However, this does not mean that any high frequency contribution is going to be preserved in the MPE-FA scheme. It will depend on its amplitude and on the shape of the density of quantization levels for that frequency.

The density of quantization levels of the different approaches for different coefficients (4, 8 and 16 cycl/deg) is shown in Figure 3. As expected from the statistical results of Figure 1, in the MSE-E and MSE-F cases, the quantization levels are more and more concentrated around zero as the frequency increases. In the MPE-FA case, the density is sharper than the cubic root of the *pdf* for low frequencies. However, the square root of the metric becomes flat as the frequency increases, i.e., the perceptual relevance of high amplitude regions increases with frequency. In the MSE-FA case, the effect of the amplitude perceptual weighting is larger in the low frequencies when $W_f(a_f)^{1/2}$ is sharper (Figure 3.a). As the metric becomes flat, the main amplitude dependency in eq. 5 is due to the *pdf* and the MSE-FA density tends to the MSE-E density.

5.3 Comparison strategy: the density surface in frequency and amplitude

The final effect of the different design criteria, MSE and MPE, is a different distribution of the total number of available reproduction levels in the encoding domain. To visualize and compare such an effect for the different approaches, the definition of a *density surface* in the frequency and amplitude plane, $\Lambda_f(a_f)$, is proposed. This surface is obtained by scaling the density of quantization levels of each coefficient by the allocated levels in that coefficient:

$$\Lambda_f(a_f) = N_f \cdot \lambda_f(a_f) \tag{13}$$

It can be interpreted as a density of quantization levels in the frequency and amplitude plane. This is simply a generalization of the role played by the the perceptual metric in the MPE approaches, for which the density surface is $\Lambda_f(a_f) \propto W_f(a_f)^{1/2}$. The density surface has the same properties as the square root of the metric in the MPE approaches:

- The location of the quantization levels of one coefficient are given by the cut of the surface for that coefficient (as in eq. 10):

$$\lambda_f(a_f) = \frac{\Lambda_f(a_f)}{\int \Lambda_f(a_f)\, da_f} \tag{14}$$

- The bit allocation is given by the integral of the density surface in amplitude (as in eq. 12):

$$b_f = \log_2 \left(\int \Lambda_f(a_f)\, da_f \right) \tag{15}$$

In this way, the relative importance given to the different regions of the encoding domain by the different design approaches is described by a single common function, $\Lambda_f(a_f)$, which completely determines the transform coder design (λ_f and b_f). The different design approaches can be understood as different ways of defining the density surface.

Figure 4 shows the density surface, $\Lambda_f(a_f)$, in MSE-E, MSE-F, MSE-FA and MPE-FA cases, along with contour views of the same functions. The integrals of the density surfaces in amplitude give the number of levels per coefficient for each approach (Figure 2).

These figures highlight the joint effect of the signal statistics and the perceptual metric in the quantizer design. In the first case, MSE-E, the density surface is fully determined by the signal statistics. The second and the third cases show the effect of the progressive introduction of frequency and amplitude perceptual features, MSE-F and MSE-FA. Figure 4.d shows the amplitude-independent surface that accounts for the simple CSF model extended for suprathreshold amplitudes. It represents the uneven bit allocation and uniform quantization of a JPEG-like approach. The last figure shows the fully perceptual MPE-FA surface: the square root of the metric. This function is inversely proportional to the contrast incremental thresholds for each frequency and amplitude (eq. 2), so it reduces to the IAF of refs. $[17, 24, 38, 41]$. In the MPE cases, the density surface is independent of the signal statistics and isolates the perceptual factors that have been progressively added in the MSE-F and MSE-FA cases.

In the MSE-F case, the frequency dependency of $W_f^{1/2}$ for low amplitudes (the CSF) cuts off the high frequency tail of the MSE-E density surface. In the MSE-FA case, the contrast non-linearities restrict the encoding effort to lower amplitudes for lower frequencies. The encoding effort is slightly extended to higher amplitudes for high frequencies following the trends of the purely perceptual function.

The differences between the MSE-FA and the MPE-FA approaches depend exclusively on the design criterion. In spite of the qualitative similarity of the density surfaces due to the introduction of the perceptual metric in the MSE approach, the actual distribution of reproduction levels is still very different. From the perceptual point of view, the encoding effort in the MSE-FA approach is still too focused on the low amplitude range, especially in high frequencies. In this way, contributions of relatively high contrast in the high frequency region will be badly represented giving rise to annoying artifacts.

5.4 Decoded images

Examples of the compression results on the standard images of Lena and Barbara are shown in Figures 5 and 6. The MSE schemes and the MPE schemes were used at the same compression ratio (0.5 bpp). These images were not included in the corresponding training process.

The experimental results show three basic trends:

- In the MSE criterion, the use of a perceptual frequency and amplitude-dependent metric, MSE-FA, does not significantly improve the final results on individual images as compared with frequency-only perceptual metrics, MSE-F, or even with euclidean metric, MSE-E.
- In the MPE criterion, the effect of the amplitude non-linearities is not negligible. As previously reported [17], the CSF-based JPEG results (MPE-F) are significantly improved if these properties are taken into account.
- The MPE-FA approach gives sharper edges and better overall subjective results, especially in the high frequency details.

These results suggest that the problems of the MSE approaches are due to the improper distribution of the quantization levels in the high frequency coefficients: they are so concentrated around zero that the less probable high amplitudes are always badly represented. This implies that the high amplitude contributions in high frequency coefficients in individual images such as Barbara's clothes and Lena's hair or eye details are severely quantized. The problem with CSF-based JPEG is its band restriction in the high frequency region in such a way that these coefficients are discarded at high compression ratios. Compare the region between 8 and 16 cycl/deg in Figures 4.f to 4.j.

The MPE-FA scheme allocates certain bits in the high frequency region but spreads the quantization levels over a relatively extensive amplitude region. In this way, the first quantization level is high enough to discard irrelevant contributions, but the levels positioned at higher amplitudes ensure an approximate representation of high contrast and high frequency contributions.

These results clarify the relative importance of the different factors affecting the quantizer design. On one hand, the relevance of the distortion metric depends on the design criterion: while the MSE approach, ruled by the signal statistics, does not exploit the eventual advantages of more sophisticated metrics, the results of the MPE approach significantly improve with the introduction of additional perceptual features in the metric. On the other hand, given an appropriate metric including frequency and amplitude features of contrast discrimination, the design criterion makes the difference: the important thing is not to minimize the average of a perceptually matched distortion, but either to ensure that, in a particular image, the MPE will be below a certain value. This bound completely depends on the perceptual geometry of the encoding domain and not on the signal statistics.

6 Final remarks

An alternative design criterion for transform coder design has been compared with the minimization of the average error, using distortion metrics that include frequency and amplitude perceptual features to different extents. The goal of the proposed criterion is to bound the maximum perceptual error in each coefficient regardless of the amplitude of the input. If a non-linear perception model and metric is assumed, this MPE criterion involves a non-uniform, input-dependent, quantization of the amplitude axis of the image transform.

Using a non-linear perceptual metric, the proposed MPE criterion improves the results of the MSE quantizers on natural imagery even if the MSE is provided with the same distortion metric. While the improvements of the proposed scheme with regard to the JPEG-like quantizer are due to the use of a more accurate perception model, its improvements with regard to the perceptually weighted MSE approach are due to the use of a perceptually meaningful design criterion. These results suggest that bounding the perceptual distortion in each particular block of the image may be more important than minimizing the *average* perceptual distortion over a set of images.

A practical side effect of the fully-perceptual MPE criterion is that it does not need off-line training stage. By its nature, it should be well-behaved over a wide range of natural images.

References

[1] N. Jayant, J. Johonston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings IEEE*, vol. 81, no. 10, pp. 1385–1422, 1993.

[2] J. Solomon, A. Watson, and A. Ahumada, "Visibility of DCT basis functions: effects of contrast masking," in *Proceedings of Data Compression Conference, Snowbird, Utah*, IEEE Computer Society Press, pp. 361–370, 1994.

[3] A. Watson, G. Yang, J. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Transactions on Image Processing*, vol. 6, pp. 1164–1175, 1997.

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

[5] A. Ahumada and H. Peterson, "Luminance-model-based DCT quantization for color image compression," vol. 1666 of *Proceedings of the SPIE*, pp. 365–374, 1992.

[6] C. Hwang, S. Venkatraman, and K. Rao, "Human visual system weighted progressive image transmission using lapped orthogonal transform classified vector quantization," *Optical Engineering*, vol. 32, no. 7, pp. 1525–1530, 1993.

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

[8] A. Leger, T. Omachi, and G. Wallace, "JPEG still picture compression algorithm," *Optical Engineering*, vol. 30, no. 7, pp. 947–954, 1991.

[9] B. Macq and H. Shi, "Perceptually weighted vector quantization in the DCT domain," *Electronics Letters*, vol. 29, no. 15, pp. 1382–1384, 1993.

[10] B. Macq, "Weighted optimum bit allocations to orthogonal transforms for picture coding," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 5, pp. 875–883, 1992.

[11] D. McLaren and D. Nguyen, "Removal of subjective redundancy from DCT-coded images," *Proceedings IEE-I*, vol. 138, no. 5, pp. 345–350, 1991.

[12] K. Nygan, K. Leong, and H. Singh, "Adaptive cosine transform coding of images in the perceptual domain," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 37, no. 11, pp. 1743–1750, 1989.

[13] K. Nygan, H. Koh, and W. Wong, "Hybrid image coding scheme incorporating human visual system characteristics," *Optical Engineering*, vol. 30, no. 7, pp. 940–946, 1991.

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

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

[16] 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.

[17] J. Malo, A. Pons, and J. Artigas, "Bit allocation algorithm for codebook design in vector quantization fully based on human visual system non-linearities for suprathreshold contrasts," *Electronics Letters*, vol. 31, no. 15, pp. 1229–1231, 1995.

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

[19] J. Lubin, "*The Use of Psychophysical Data and Models in the Analysis of Display System Performance*," in *Digital Images and Human Vision* (A. Watson, ed.), (Massachusetts), pp. 163–178, MIT Press, 1993.

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

[21] 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.

[22] H. Wilson, "Pattern discrimination, visual filters and spatial sampling irregularities," in *Computational Models of Visual Processing* (M. Landy and J. Movshon, eds.), (Massachussets), pp. 153–168, MIT Press, 1991.

[23] A. Pons, J. Malo, J. Artigas, and P. Capilla, "Image quality metric based on multidimensional contrast perception models," *Displays*, Accepted Feb. 1999.

[24] 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 and Vision Computing*, vol. 15, no. 7, pp. 535–548, 1997.

[25] D. Kelly, "Receptive field like functions inferred from large area psychophysical measurements," *Vision Research*, vol. 25, no. 12, pp. 1895–1900, 1985.

[26] N. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Transactions on Communications*, vol. 33, pp. 551–557, 1985.

[27] L. Saghri, P. Cheatheam, and A. Habibi, "Image quality measure based on a human visual system model," *Optical Engineering*, vol. 28, no. 7, pp. 813–819, 1989.

[28] J. Daugman, "Uncertainity relation for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America*, vol. 2, no. 7, pp. 1160–1169, 1985.

[29] S. Marcelja, "Mathematical description of the response of simple cortical cells," *Journal of the Optical Society of America*, vol. 70, no. 11, pp. 1297–1300, 1980.

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

[31] J. Malo, A. Pons, A. Felipe, and J. Artigas, "Characterization of human visual system threshold performance by a weighting function in the Gabor domain," *Journal of Modern Optics*, vol. 44, no. 1, pp. 127–148, 1997.

[32] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory*, vol. 14, pp. 676–683, 1968.

[33] Y. Yamada, S. Tazaki, and R. Gray, "Asymptotic performance of block quantizers with difference distortion measures," *IEEE Transactions on Information Theory*, vol. 26, no. 1, pp. 6–14, 1980.

[34] W. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, 1948.

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

[36] A. Gersho, "Asymptotically optimal block quantization," *IEEE Transactions on Information Theory*, vol. 25, no. 4, pp. 373–380, 1979.

[37] J. Huang and P. Schultheiss, "Block quantization of correlated gaussian random variables," *IEEE Transactions on Communications Systems*, vol. 11, no. 3, pp. 289–296, 1963.

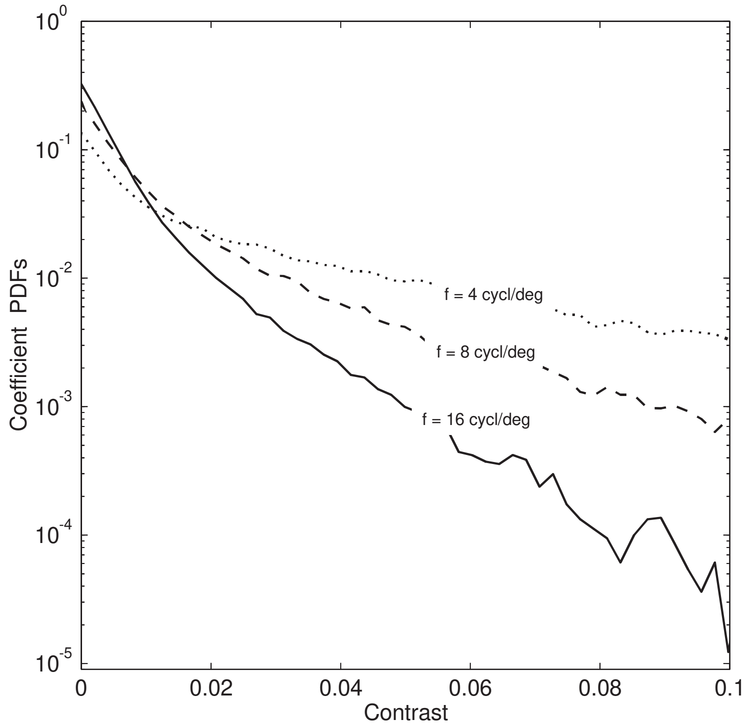
[38] J. Malo, F. Ferri, J. Albert, and J. Artigas, "Adaptive motion estimation and video vector quantization based on spatio-temporal non-linearities of human perception," *Lecture Notes on Computer Science, Springer Verlag*, vol. 1310, pp. 454–461, 1997.

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

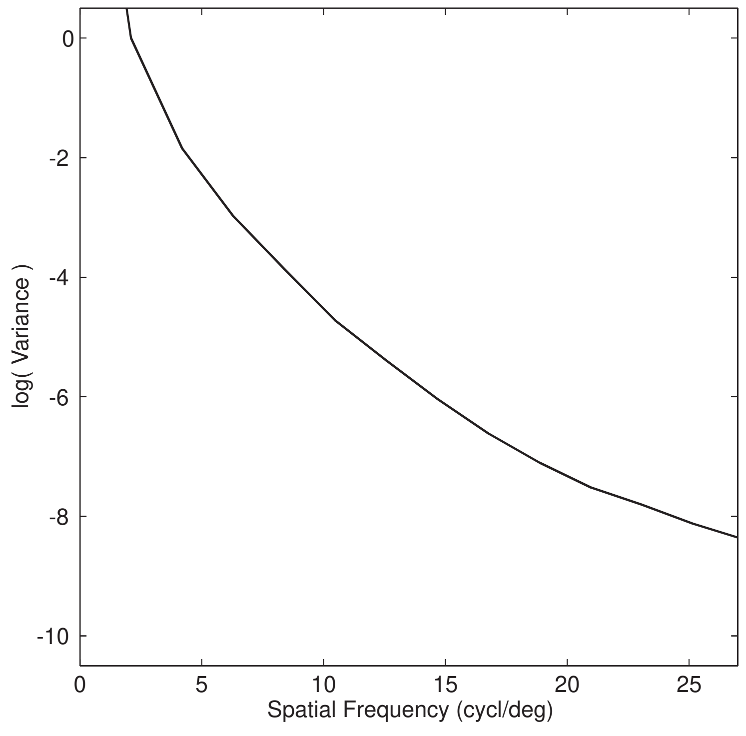
[40] S. Daly, "Visible differences predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision* (A. Watson, ed.), (Massachusetts), pp. 179–206, MIT Press, 1993.

[41] J. Malo, F. Ferri, J. Albert, and J. Artigas, "Splitting criterion for hierarchical motion estimation based on perceptual coding," *Electronics Letters*, vol. 34, no. 6, pp. 541–543, 1998.

[42] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, 1980.

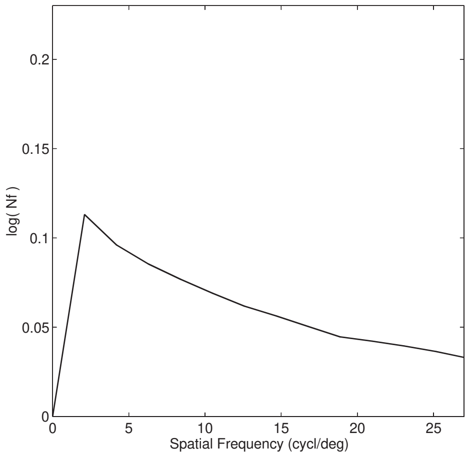


a)

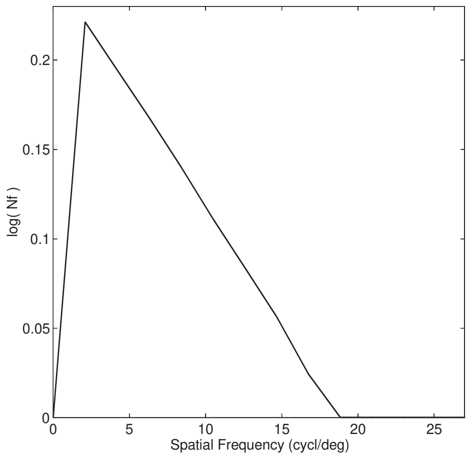


b)

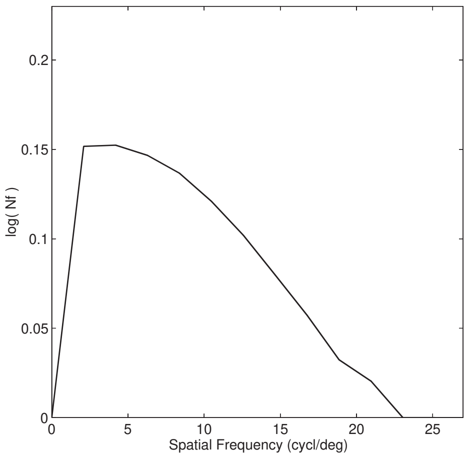
Fig. 1. a) Contrast (normalized amplitude) pdfs for some transform coefficients, b) Variance of the coefficients in the training set.



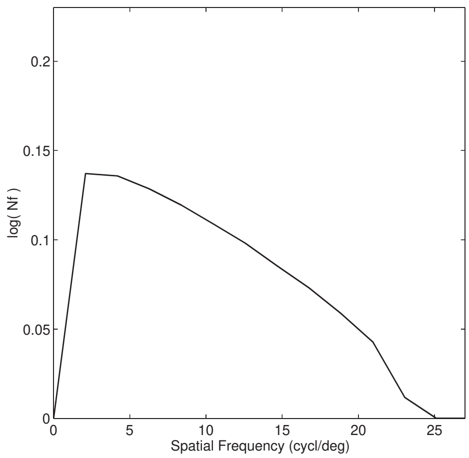
a)



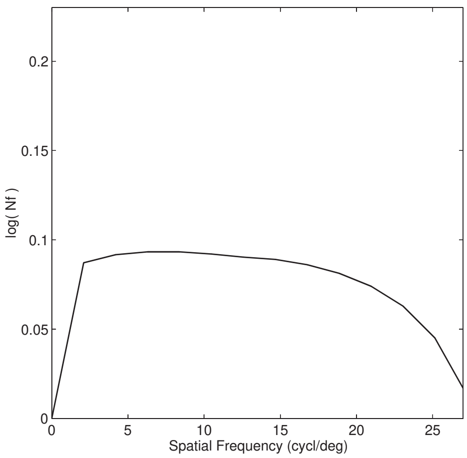
b)



c)

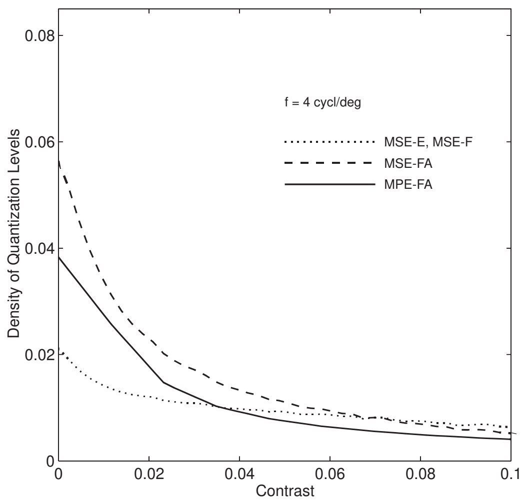


d)

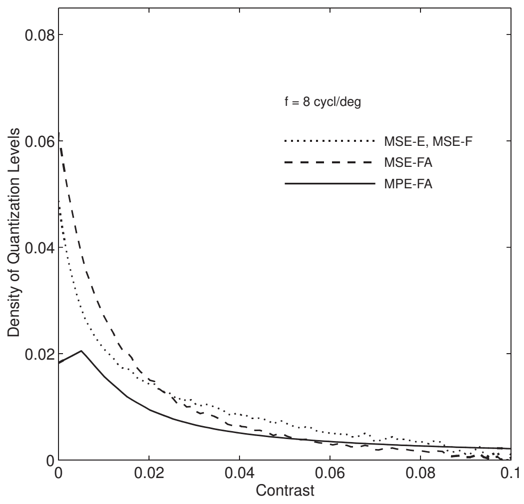


e)

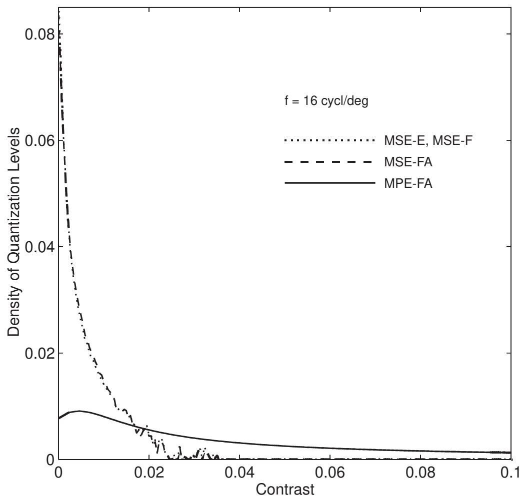
Fig. 2. Bit allocation curves (log of the allocated levels). a) MSE-E, b) MSE-F, c) MSE-FA, d) MPE-F, e) MPE-FA. The curves are set to zero for the first coefficient, the DC component, because it is DPCM coded in every approach.



a)



b)



c)

Fig. 3. Densities of quantization levels, $\lambda_f(a_f)$, for different coefficients using different design criteria. a) 4 cycl/deg, b) 8 cycl/deg, c) 16 cycl/deg.

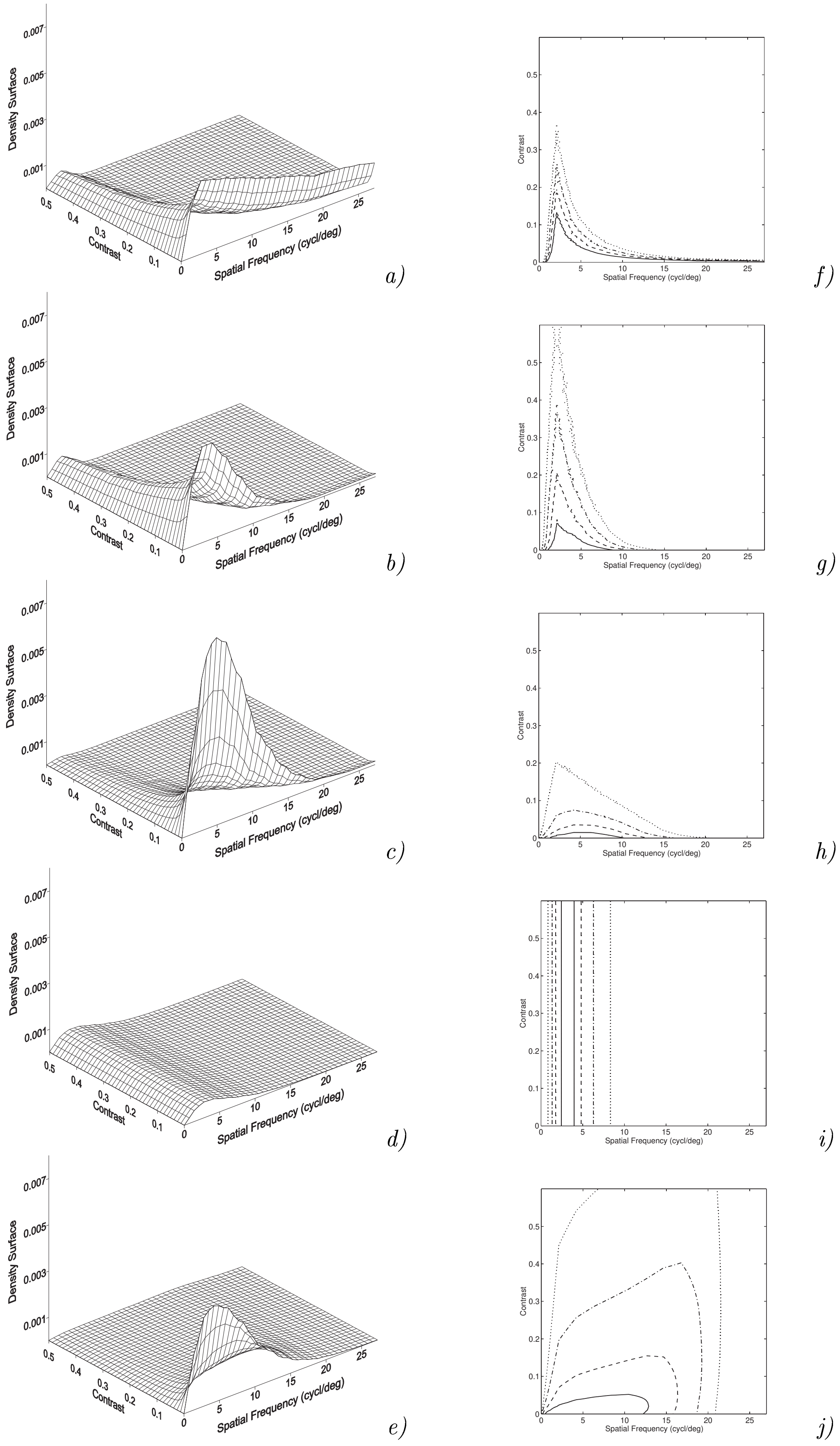


Fig. 4. Density surfaces (a-e) and their corresponding contour plots (f-j) for the different approaches. From top to bottom, MSE-E, MSE-F, MSE-FA, MPE-F and MPE-FA. The contour lines enclose the encoding regions with the 80%, 60%, 40% and 20% of the quantization levels. They show where the encoding effort is focused in the different approaches.

Fig. 5. *Compression results on Barbara image (0.5 bpp). The figures show the* 170×170 *central subimage.*

Fig. 6. *Compression results on Lena image (0.5 bpp). The figures show the* 170×170 *central subimage.*