

Lec. 5 Optimal scalar quantization of images in signal and transform domains

5.1 Principles of optimal scalar quantization

Scalar (element-wise) quantization is the second stage of signal digitization. It is applied to signal discrete representation obtained as the result of signal discretization. Scalar quantization implies that a dynamic range of a finite length is chosen in the entire range of the continuous signal representation coefficient values and is divided into *quantization intervals* (Fig. 5-1). Any value falling within a particular interval is designated by a number (quantization interval's index) common to all values in the interval. In signal reconstruction, this number is replaced with a value selected as a representative of this interval (*quantized value*).

Arrangement of quantization intervals and selection of representative values are governed by the accuracy criterion for representing the continuous signal by a digital one and by probability distribution of the signal values.

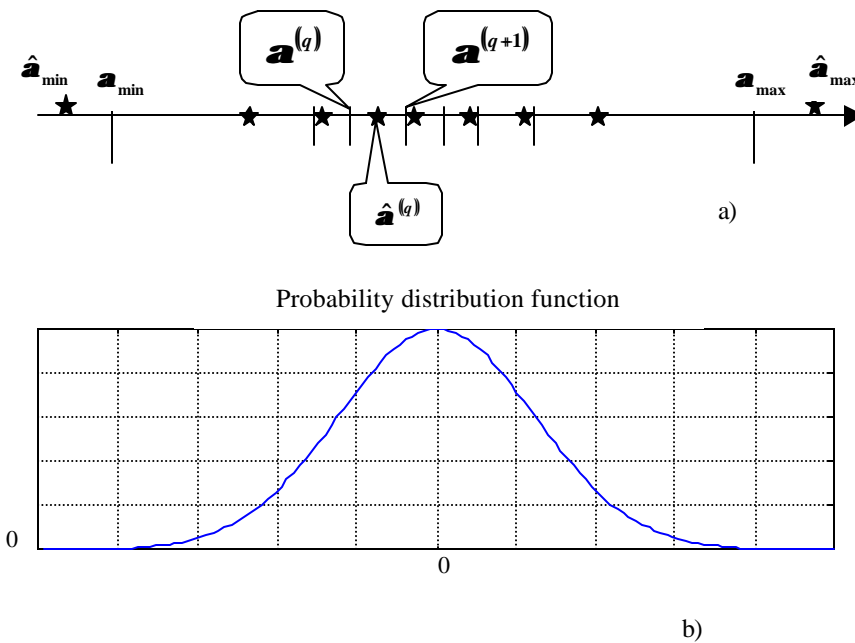


Figure 5-1. Non-uniform quantization. (a)-arrangement of quantization intervals and their representative values; (b) - probability distribution function of values to be quantized; (c) - non-uniform quantization by means of a nonlinear pre-distortion

Let \mathbf{a} be a value of signal's discrete representation coefficient to be quantized, $\hat{\mathbf{a}}_{\min}$ and $\hat{\mathbf{a}}_{\max}$ are the corresponding representatives of values outside the selected dynamic range, \mathbf{a}_{\min} and \mathbf{a}_{\max} are boundaries (minimum and maximum) of the dynamic range, $\mathbf{a}^{(q)}$, $\mathbf{a}^{(q+1)}$ and $\hat{\mathbf{a}}^{(q)}$, are, respectively, the left and right boundaries and representative value of the q -th quantization interval (Fig. 5-1, a). One can characterize the quantization error at the q -th interval within the dynamic range as

$$\mathbf{e}^{(q)} = \mathbf{a} - \hat{\mathbf{a}}^{(q)}; \mathbf{a} \in [\mathbf{a}^{(q)}, \mathbf{a}^{(q+1)}] \quad (5.1)$$

and truncation errors of the dynamic range limitation as

$$\mathbf{e}^{(\min)} = \mathbf{a} - \hat{\mathbf{a}}^{(\min)}; \mathbf{a} < \mathbf{a}_{\min}; \quad (5.2)$$

$$\mathbf{e}^{(\max)} = \mathbf{a} - \hat{\mathbf{a}}^{(\max)}; \mathbf{a} > \mathbf{a}_{\max}. \quad (5.3)$$

The requirements for quantization accuracy are generally formulated in terms of the constraints imposed upon the quantization errors $\{\mathbf{e}^{(q)}\}$. The most common approach to formulating these constraints assumes that the quantization artifacts are the same for all signal representation coefficients and therefore all coefficients are quantized in the same way. Such quantization is called *homogeneous*.

For the design of homogeneous quantizer, quantization artifacts are evaluated on average over all possible coefficient values to be quantized. To this goal, loss functions $D_l(\mathbf{e}^{(\min)})$, $D_{dr}(\mathbf{e}^{(q)})$, $D_r(\mathbf{e}^{(q)})$ that measure losses owing to the quantization errors within and outside the dynamic range and probability density $p(\mathbf{a})$ of the values to be quantized are introduced, and quantization quality is evaluated, separately for quantization errors within and outside the dynamic range, as

$$\bar{D}_l = \int_{\mathbf{a}_{\min}}^{\mathbf{a}_{\max}} p(\mathbf{a}) D_l(\mathbf{e}_{\min}) d\mathbf{a} \quad (5.4)$$

$$\bar{D}_r = \int_{\mathbf{a}_{\max}}^{\mathbf{a}_{\max}} p(\mathbf{a}) D_r(\mathbf{e}_{\max}) d\mathbf{a} \quad (5.5)$$

and

$$\bar{D}_{dr} = \sum_{q=1}^{Q-1} \int_{\mathbf{a}^{(q)}}^{\mathbf{a}^{(q+1)}} p(\mathbf{a}) D_{dr}(\mathbf{e}^{(q)}) d\mathbf{a}; \mathbf{a}^{(1)} = \mathbf{a}_{\min}; \mathbf{a}^{(Q)} = \mathbf{a}_{\max}, \quad (5.6)$$

where Q is the number of quantization intervals. Separate evaluation of dynamic range limitation errors and quantization errors is advisable owing to the different nature of these errors: while quantization errors are limited in the range by the quantization interval size, dynamic range limitation errors may be substantially larger.

Eqs. 5.4 and 5 can be used for determining dynamic range boundaries $\{\mathbf{a}_{\min}, \mathbf{a}_{\max}\}$ given dynamic range limitation errors \bar{D}_l and \bar{D}_r . Eq. (5.6) can be used for determining the set of quantization intervals boundaries $\{\mathbf{a}^{(q)}\}$ and representatives $\{\hat{\mathbf{a}}^{(q)}\}$ that minimize the number of quantization intervals Q given average quantization error \bar{D}_{dr} or minimize the average quantization error given the number of quantization intervals. Such set of quantization interval boundaries and quantization interval representatives determine *optimal scalar quantizer*.

Optimal quantization intervals are, in general, non-uniform and are defined by two factors: by probability density $p(\mathbf{a})$ and by loss function $D_{dr}(\mathbf{e})$. The lower is the probability density in a certain sub-range of the dynamic range of \mathbf{a} the smaller is contribution of the quantization error in this sub-range into the average quantization error and therefore the larger can be quantization intervals in this sub-range. The loss function affects size of the quantization intervals similarly.

5.2. Implementations of optimal non-uniform quantization

There are two approaches to the design of the optimal quantizer, a numerical optimization approach and a *compressor-expander (compander)* one. Numerical optimization approach assumes analytical or numerical solving the optimization equation:

$$\{\mathbf{a}^{(q)}, \hat{\mathbf{a}}^{(q)}\} = \arg \min_{\{\mathbf{a}^{(q)}, \hat{\mathbf{a}}^{(q)}\}} \bar{D}_{dr} = \sum_{q=1}^{Q-1} \int_{\mathbf{a}^{(q)}}^{\mathbf{a}^{(q+1)}} p(\mathbf{a}) D_{dr}(\mathbf{e}^{(q)}) d\mathbf{a}. \quad (5.7)$$

The optimization is much simplified if loss function $D_{dr}(\mathbf{e}^{(q)})$ is an even function: $D_{dr}(\mathbf{e}) = D_{dr}(-\mathbf{e})$ such as, for instance, a quadratic loss function $D_{dr}(\mathbf{e}) = \mathbf{e}^2$ is. In this case, from $\frac{1}{\mathbf{1}\mathbf{a}^{(q)}} D_{dr} = 0$, it follows that optimal boundaries of the discretization intervals should be places just halfway from the corresponding quantized values:

$$\mathbf{a}_{opt}^{(q)} = (\hat{\mathbf{a}}_{opt}^{(q-1)} + \hat{\mathbf{a}}_{opt}^{(q)}) / 2. \quad (5.8)$$

As for the representatives, they still should be found by a numerical or, when it is possible, by analytical optimization. For instance, for the quadratic loss function, from $\frac{1}{\mathbf{1}\hat{\mathbf{a}}^{(q)}} D_{dr} = 0$ it follows, in

addition to Eq. (5.8), that optimal representatives are centers of mass of the probability density within the discretization intervals:

$$\hat{\mathbf{a}}_{opt}^{(q)} = \frac{\int_{\mathbf{a}^{(q)}}^{\mathbf{a}^{(q+1)}} \mathbf{a} p(\mathbf{a}) d\mathbf{a}}{\int_{\mathbf{a}^{(q)}}^{\mathbf{a}^{(q+1)}} p(\mathbf{a}) d\mathbf{a}} \quad (5.9)$$

Such a solution of the quantization optimization problem for the quadratic loss function is called **Max-Lloyd quantization**.

Compressor-expander quantization was initially suggested as a method for hardware implementation of non-uniform optimal quantization using readily available uniform quantizers in which signal dynamic range is split into equal quantization intervals and centers of the intervals are used as quantization levels. In order to implement a non-uniform quantization with uniform quantizers, signal, before being sent to uniform quantizer is subjected to a compressive nonlinear point wise transformation. Correspondingly, at the signal reconstruction stage, quantized values are to be subjected, in digital-to-analog converters, to the expanding nonlinear transformation inverse to the compressing one. (Fig. 5-2).

Optimization of the compressor-expander quantization is achieved by an appropriate selection of the compressive nonlinear point wise transformation. For analytical optimization, quantization optimization equation (5.7) should be reformulated in terms of the compression transformation function. Let $w(\cdot)$ is a compression transformation function, D_u is interval of uniform quantization.

Then, for a particular value \mathbf{a} to be quantized, quantization interval D that corresponds to uniform quantization of values of function $w(\mathbf{a})$ will be equal to (Fig. 5-2, a)

$$D = \frac{D_u}{dw(\mathbf{a})/d\mathbf{a}} \quad (5.10)$$

and one can obtain optimal function $w(\mathbf{a})$ as a solution of equation:

$$w(\mathbf{a}) = \arg \min_{w(\mathbf{a})} \int_{\mathbf{a}_{min}}^{\mathbf{a}_{max}} p(\mathbf{a}) D_{dr} \left(\frac{D_u}{dw(\mathbf{a})/d\mathbf{a}} \right) d\mathbf{a}, \quad (5.11)$$

where $D_{dr}(\cdot)$ is a quantization loss function formulated in terms of the range of quantization errors for each particular value \mathbf{a} . The solution is provided by the Euler-Lagrange equation, which in this case is written as:

$$\frac{\int_{\mathbf{a}_{min}}^{\mathbf{a}_{max}} p(\mathbf{a}) D_{dr} \left(\frac{D_u}{dw(\mathbf{a})/d\mathbf{a}} \right) d\mathbf{a}}{\int_{\mathbf{a}_{min}}^{\mathbf{a}_{max}} p(\mathbf{a}) D_{dr} \left(\frac{D_u}{dw(\mathbf{a})/d\mathbf{a}} \right) d\mathbf{a}} = const, \quad (5.12)$$

where $\dot{w} = dw(\mathbf{a})/d\mathbf{a}$.

The following special cases will provide an insight into how probability distribution $p(\mathbf{a})$ and the type of loss function $D_{dr}(\cdot)$ affect optimal arrangement of quantization intervals within the quantized value dynamic range. We will begin with threshold quantization quality criteria that assume that quantization error is nil if it does not exceed a certain threshold. For threshold criteria, optimal arrangement of quantization intervals does not depend on probability distribution $p(\mathbf{a})$ because, according to the criteria, losses owing to quantization should be kept zero for all quantization intervals.

5.3 Examples of optimal scalar compader quantization

Example 1. Uniform threshold criterion for Absolute value of Quantization Error:

$$\bar{D}_{dr} \left(\frac{D_u}{\dot{w}} \right) = \begin{cases} 0, & |D| \leq D_{thr} \\ 1, & \text{otherwise} \end{cases} \quad (5.13)$$

From Eq. (5.13) it follows that, for monotonic functions $w(\mathbf{a})$,

$$\dot{w} = 2D_u / D_{thr}, \quad (5.14)$$

and, therefore, uniform quantization is optimal:

$$\frac{w(\mathbf{a}) - w(\mathbf{a}_{min})}{w(\mathbf{a}_{max}) - w(\mathbf{a}_{min})} = \frac{\mathbf{a} - \mathbf{a}_{min}}{\mathbf{a}_{max} - \mathbf{a}_{min}}. \quad (5.15)$$

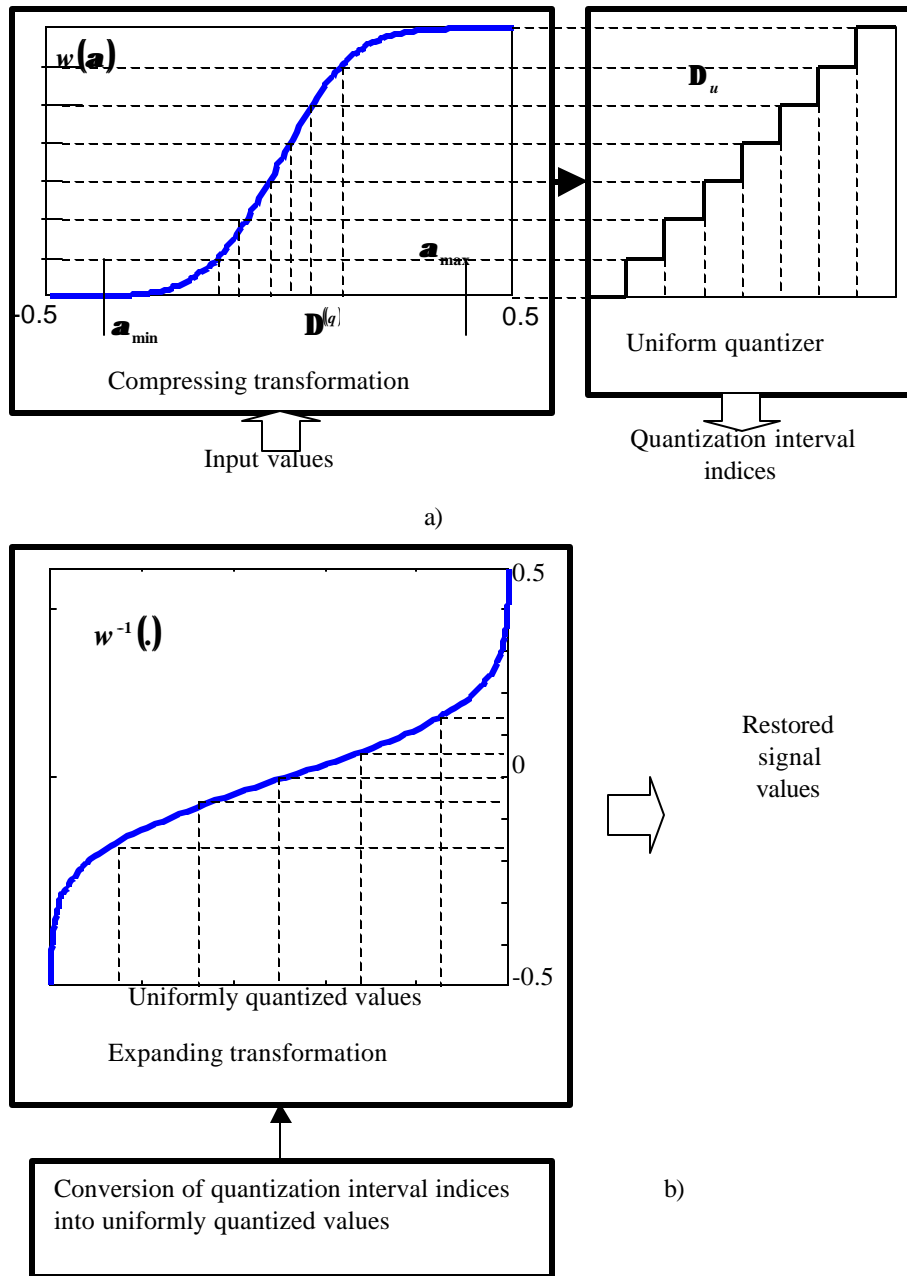


Figure 5-2. Schematic diagram of the companding quantization (a) and expanding restoration (b)

Example 2. Uniform threshold criterion of Relative value of Quantization Error:

$$\bar{D}_{dr} \begin{cases} \frac{2D_u}{\bar{a}} / \frac{\bar{a}}{w} = \frac{\bar{a}}{w} > 1, & |D| \leq D_{thr} = d_{thr} \bar{a} \\ \frac{2D_u}{\bar{a}} / \frac{\bar{a}}{w} = \frac{\bar{a}}{w} \leq 1, & \text{otherwise} \end{cases} \quad (5.16)$$

In this case,

$$\bar{w} = 2D_u / d_{thr} \bar{a}, \quad (5.17)$$

and, therefore, uniform quantization in a logarithmic scale is optimal:

$$\frac{w(\bar{a}) - w(\bar{a}_{min})}{w(\bar{a}_{max}) - w(\bar{a}_{min})} = \frac{\ln(\bar{a}/\bar{a}_{min})}{\ln(\bar{a}_{max}/\bar{a}_{min})}. \quad (5.18)$$

This particular case is of a special interest in image processing. Signal quantization results in piece wise constant signals. In quantized images, boundaries between these pieces visually appear as “*false contours*” (see Fig. 5-3).

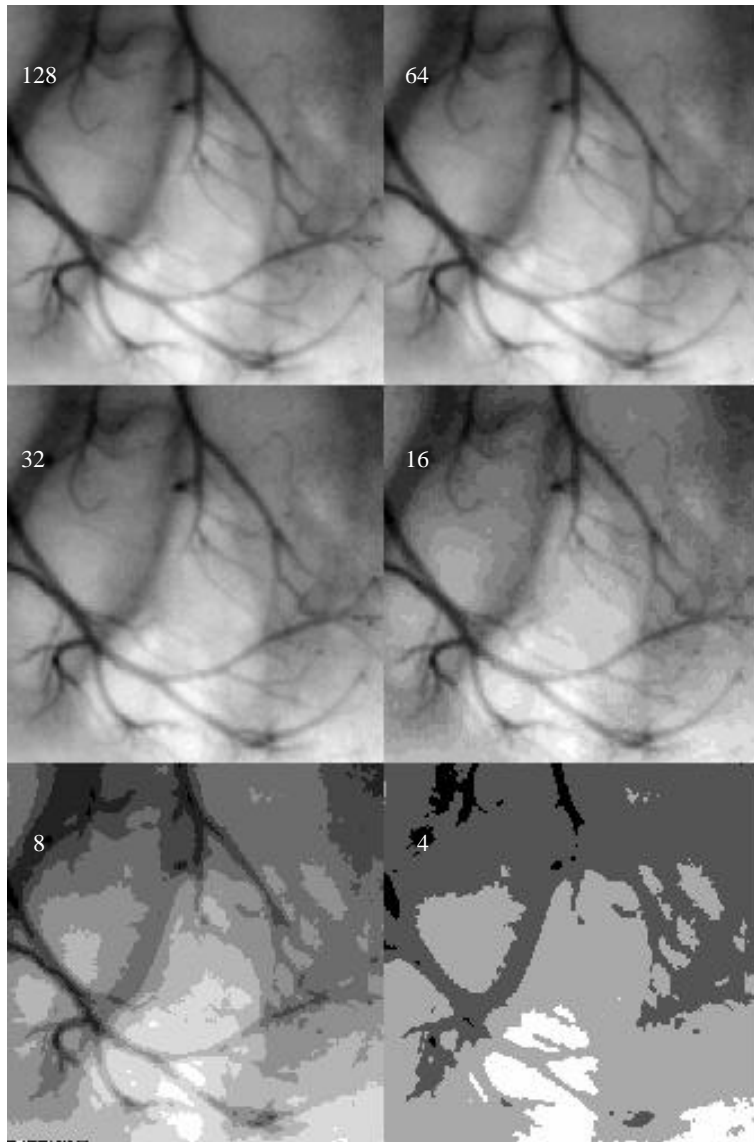


Figure 5-3. False contours in image quantization (128; 64; 32; 16; 8; 4 quantization levels)

A natural primary requirement for image quantization is that the number and arrangement of quantization levels should be selected so as to secure invisibility of false contours in displayed digital images. Visibility of patches of constant gray level depends on their contrast and of size of patches the constant gray level. According to known *Weber-Fechner's law*, the ratio of brightness contrast \mathbf{DB}_{thr} of a stimulus with brightness $B + \mathbf{DB}_{thr}$ on the threshold of its visibility to background brightness B is constant for a wide range of the brightness (Fig. 5-3):

$$\frac{\mathbf{DB}_{thr}}{B} = \mathbf{a}_{thr} = const . \quad (5.19)$$

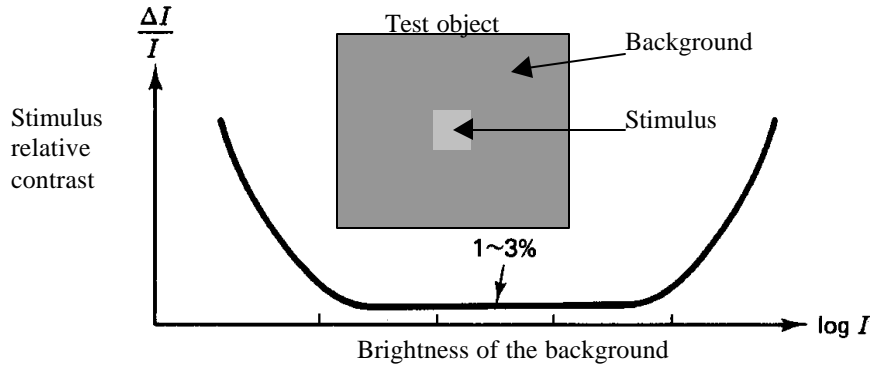


Fig. 5-3 Sensitivity curve of human vision

The constant decreases with the stimulus size. Its lowest value of 1% - 3% corresponds to stimuli of large angular size. Therefore it follows from Weber-Fechner's law that optimal, for digital image display, quantization of image samples should be uniform in the logarithmic scale. The number Q of required quantization levels for logarithmic quantization can be found from equation

$$Q = \frac{w(\mathbf{a}_{\max}) - w(\mathbf{a}_{\min})}{D_u} = \frac{\ln(\mathbf{a}_{\max} / \mathbf{a}_{\min})}{d_{hr}} \quad (5.20)$$

For good quality photographic, TV and computer displays, dynamic range $\mathbf{a}_{\max} / \mathbf{a}_{\min}$ of displayed image brightness is about 100. Using this estimation of the dynamic range and taking visual sensitivity threshold equal to 2%, one can obtain from Eq. 5.20 that the required number of quantization levels is 243. Similar figures are characteristic also for human audio sensibility. This was the main reason why 256 quantization levels (8 bits) were chosen as a standard for image, sound and other analog signal representation and why 8 bits (byte) had become a basic unit in the computer industry.

Example 3. Exponential Criterion of Absolute Quantization Error:

$$\bar{D}_{dr}(\mathbf{D}) = |\mathbf{D}|^{2p} = \left| \frac{D_u}{w} \right|^{2q} \quad (5.21)$$

Substituting (5.21) into (5.12) and solving the resulting differential equation, obtain

$$\frac{w(\mathbf{a}) - w(\mathbf{a}_{\min})}{w(\mathbf{a}_{\max}) - w(\mathbf{a}_{\min})} = \frac{\int_{\mathbf{a}_{\min}}^{\mathbf{a}} \dot{p}(\mathbf{a})^{1/(2q+1)} d\mathbf{a}}{\int_{\mathbf{a}_{\min}}^{\mathbf{a}_{\max}} \dot{p}(\mathbf{a})^{1/(2q+1)} d\mathbf{a}} \quad (5.22)$$

Thus, the required degree of nonlinear pre-distortion depends solely on the probability distribution of the quantized values. The meaning of this relationship becomes evident from the expression:

$$\mathbf{D}(\mathbf{a}) = D_u / w(\mathbf{a}) \mu(p(\mathbf{a}))^{-1/(2q+1)} \quad (5.23)$$

which implies that the size of quantization intervals for the various values of \mathbf{a} are inversely proportional to their probability densities raised to the corresponding power. For the widely used mean squared quantization error criterion ($q = 1$),

$$\frac{w(\mathbf{a}) - w(\mathbf{a}_{\min})}{w(\mathbf{a}_{\max}) - w(\mathbf{a}_{\min})} = \frac{\int_{\mathbf{a}_{\min}}^{\mathbf{a}} \dot{p}(\mathbf{a})^{1/3} d\mathbf{a}}{\int_{\mathbf{a}_{\min}}^{\mathbf{a}_{\max}} \dot{p}(\mathbf{a})^{1/3} d\mathbf{a}} \quad (5.24)$$

Note that an interesting special case

$$\frac{w(\mathbf{a}) - w(\mathbf{a}_{\min})}{w(\mathbf{a}_{\max}) - w(\mathbf{a}_{\min})} = \frac{\int_{\mathbf{a}_{\min}}^{\mathbf{a}} \dot{p}(\mathbf{a}) d\mathbf{a}}{\int_{\mathbf{a}_{\min}}^{\mathbf{a}_{\max}} \dot{p}(\mathbf{a}) d\mathbf{a}} \quad (5.25)$$

takes place when $q = 0$. Such a transformation is known as *histogram equalization* and is one of popular image enhancement transformations. For image quantization, this transformation assumes that

quantization errors are equally important whatever they values are (see Eq. (5.21)) and makes quantization intervals to be inversely proportional to the probability density for the level to be quantized (see Eq. 5.23). Sometimes, as in the case quantizing spectral coefficients of signals in Fourier, Walsh and other bases, one may regard the quantized coefficients of the discrete signal representation as distributed according to a truncated Gaussian probability density distribution on the interval $[a_{\min}, a_{\max}]$:

$$p(a) \propto \exp\left[-\frac{(a - \bar{a})^2}{2s_a^2}\right]. \quad (5.26)$$

Then, for mean squared error criterion ($P = 1$),

$$\frac{w(a) - w(a_{\min})}{w(a_{\max}) - w(a_{\min})} = \frac{F\left[\frac{(a - \bar{a})}{\sqrt{3}s_a}\right] - F\left[\frac{(a_{\min} - \bar{a})}{\sqrt{3}s_a}\right]}{F\left[\frac{(a_{\max} - \bar{a})}{\sqrt{3}s_a}\right] - F\left[\frac{(a_{\min} - \bar{a})}{\sqrt{3}s_a}\right]} \quad (5.27)$$

where

$$F(x) = \int_{-\infty}^x \exp\left(-\frac{x^2}{2}\right) dx. \quad (5.28)$$

One can find that, for this case, the reduction in the required number of quantization levels, as compared, for the same averaged quantization error, to the case of uniform quantization, is equal to

$$g = \frac{a_{\max/\min}}{\sqrt[4]{27} \sqrt{2} P} \frac{[F(a_{\max/\min}/2) - F(a_{\max/\min}/2)]^{1/2}}{[F(a_{\max/\min}/2\sqrt{3}) - F(a_{\max/\min}/2\sqrt{3})]^{1/2}}, \quad (5.29)$$

where $a_{\max/\min} = (a_{\max} - a_{\min})/s_a$. For large $a_{\max/\min}$, the gain tends to $a_{\max/\min}/5.7$.

Example 4. Exponential Criterion of Relative Quantization Error:

$$\bar{D}_{dr}(D) = |D/a|^{2P} = \left| \frac{D}{w(a)} \right|^{2q} \quad (5.30)$$

In this case, solution of the Euler-Lagrange equation (5.12) yields:

$$\frac{w(a) - w(a_{\min})}{w(a_{\max}) - w(a_{\min})} = \frac{\int_{a_{\min}}^a (p(a)/a^{2q})^{1/(2q+1)} da}{\int_{a_{\min}}^{a_{\max}} (p(a)/a^{2q})^{1/(2q+1)} da}. \quad (5.31)$$

If quantized values a are distributed uniformly in the dynamic range, optimal compression function is

$$\frac{w(a) - w(a_{\min})}{w(a_{\max}) - w(a_{\min})} = \frac{a^{1/(2q+1)} - a_{\min}^{1/(2q+1)}}{a_{\max}^{1/(2q+1)} - a_{\min}^{1/(2q+1)}}. \quad (5.32)$$

For the mean squared relative quantization error criterion (Eq. (5.30), $q = 1$):

$$\frac{w(a) - w(a_{\min})}{w(a_{\max}) - w(a_{\min})} = \frac{a^{1/3} - a_{\min}^{1/3}}{a_{\max}^{1/3} - a_{\min}^{1/3}}. \quad (5.33)$$

We will refer to such type of nonlinear transformations of Eq. (5.32) as to "*P-th law quantization*":

$$\frac{w(a) - w(a_{\min})}{w(a_{\max}) - w(a_{\min})} = \frac{a^P - a_{\min}^P}{a_{\max}^P - a_{\min}^P}. \quad (5.34)$$

P-th law quantization proved to be very useful for quantizing Fourier and DCT transform coefficients of images. The optimal value of the nonlinearity index *P* for which standard deviation of errors in reconstructed images because of quantization of their spectra is minimal is usually about 0.3 (see, for instance, Fig. 5-5).

Remarkable enough, this corresponds to the optimum for the mean squared relative quantization error criterion defined by Eq. 5.3 3. Another option in quantization image spectral coefficients is *inhomogeneous quantization* in which different coefficients are quantized in different way. Inhomogeneous quantization in for of *zonal quantization* has found its application in image transform coding.

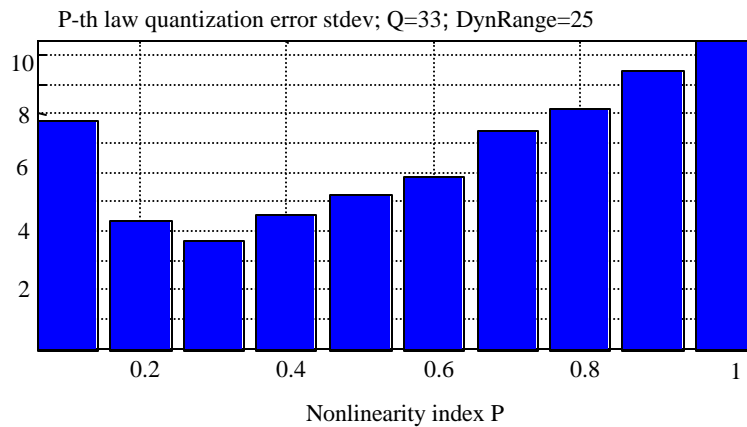
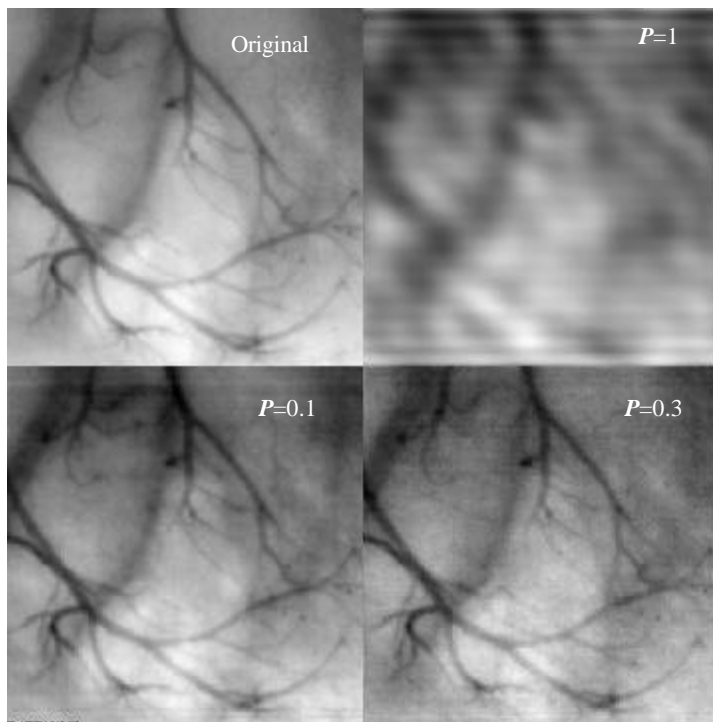


Figure 5-5. Optimization of P -th law quantization of image spectrum (33 quantization levels of spectrum real and imaginary parts in the dynamic range of ± 25 of their standard deviation): Initial image and images reconstructed from uniformly ($P=1$) and P -th ($P=0.3$ and 0.1) law quantized of spectral coefficients. Bar diagram at the bottom shows how standard deviation of quantization error depends on the nonlinearity index P . Note that while not much difference between images for $P=0.1$ and $P=0.3$ is visible, standard deviations of the quantization error differ substantially for these two cases.