RESOLUTION/QUANTIZATION TRADE-OFFS IN IMAGE/SIGNAL REPRESENTATION

Alfred M Bruckstein

Jury Lecture

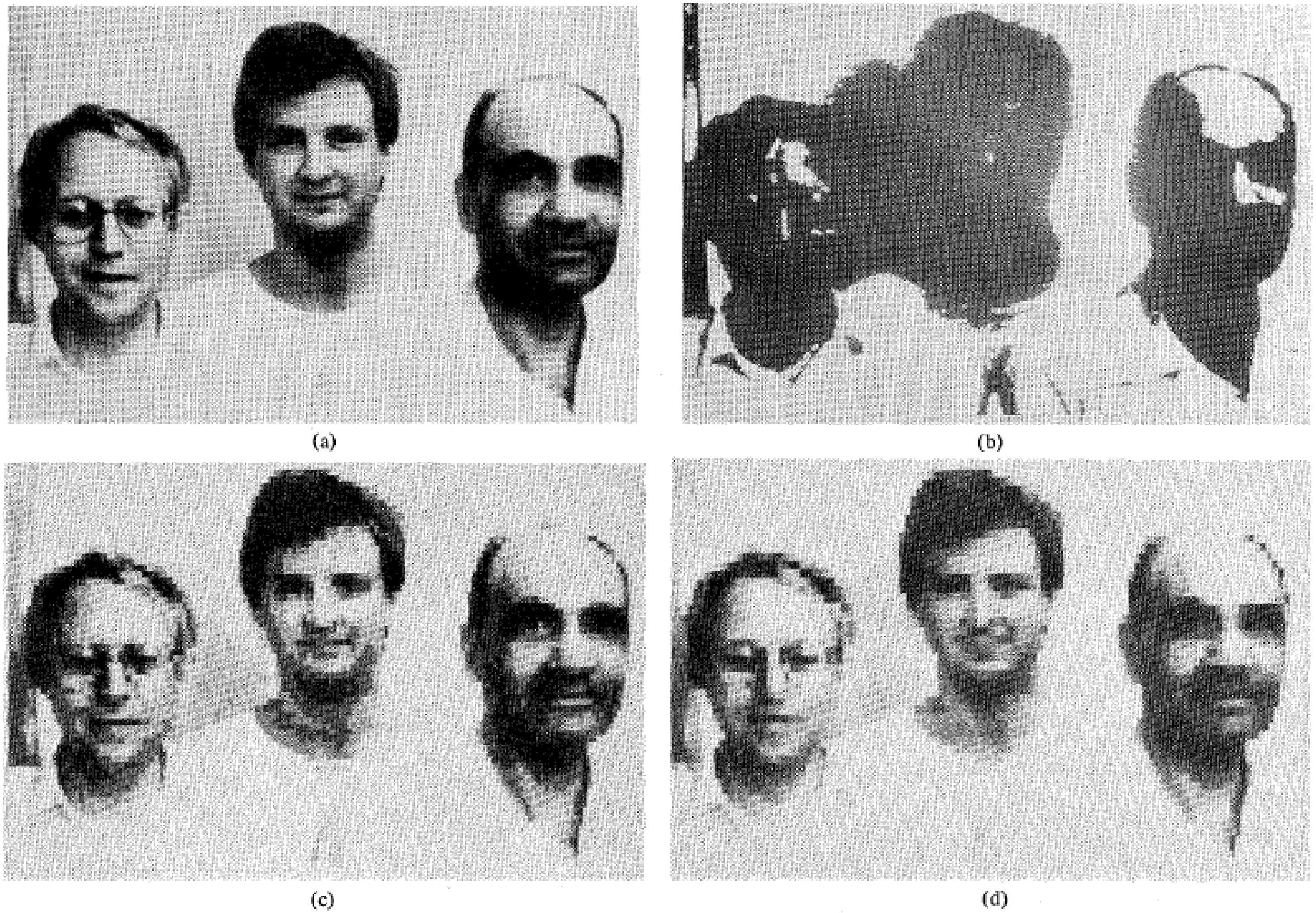


Fig. 1. Original image of the authors (a) and digitized images (b) $M = 256$, $N = 256$, $b = 1$; (c) $M = 170$, $N = 85$, $b = 4$; (d) $M = 128$, $N = 64$, $b = 7$. The images require (b) 65 536, (c) 57 800, and (d) 57 344 bits. The resulting normalized values of the criterion are (b) 6.56, (c) 1, and (d) 1.59. The image (c) is closest to the optimal.

Optimal Digitization of 2-D Images

L. NIELSEN, K. J. ÅSTRÖM, AND E. I. JURY

Abstract—The problem of representing an image by $M \times N$ samples with b bits/sample, subject to the constraint of a fixed total number of bits, is discussed. Reasonable assumptions are made in such a way that the optimization problem has a closed form solution. The solution is tested experimentally and agrees well with human perception of visual quality. The analytical solution brings out the dependence of the optimal digitization on image characteristics very clearly. The results explain and agree with results of other subjective tests.

I. Introduction

Vigorous research has been devoted to image processing and related fields during the last two decades. Several books have been written on various aspects of the theory and applications [1]–[3]. The problem of optimal digitization of 2-D images has been sporadically mentioned in several texts, but it has not been addressed in full detail [2]–[4]. Experimental investigations of the effect of coarse scan/fine print for bilevel images was initiated by Abdou and Wong [5]. No theory was given in this work. Steiglitz [6] has presented a detailed theory of transmission of an anlog signal over a fixed bit-rate channel in the 1-D case. This work has motivated the extension to the 2-D case in this paper, which gives a definition of optimal digitization (quantization and sampling) of images for the first time. The definition seems to be meaningful in practical applications. It differs from Steiglitz work in the respect that our problem formulation admits an analytical solution.

II. Problem Formulation

Let the original image $f(x, y)$ be a function defined on $\Omega = [0, L_x] \times [0, L_y] \subset R^2$ with values in $V = [\min f, \max f] \subset R^+$. The image is sampled to give a sampled image $\hat{f}(x_i, y_j)$ defined on an $M \times N$ rectangular grid G with values in $V \subset R^+$. Ideal sampling is assumed, i.e., $\hat{f}(x_i, y_j) = f(x_i, y_j)$ and $x_i, y_j \in G$. In addition to sampling, the values of $\hat{f}(x_i, y_j)$ are also quantized so that $V \subset R^+$ is represented by b bits in 2^b quantization levels. The quantization of $\hat{f}$ is denoted by $Q\hat{f}$, which is the digitized image defined on an $M \times N$ grid with discrete values. We want to reconstruct the new function $\tilde{f}$ defined on Ω with values in V. From the function $Q\hat{f}$ the function $\tilde{f}$ can be obtained using many different interpolation schemes [1], [5], [6].

The optimal digitization problem can be formulated as follows. Assume that the image is represented by a fixed number of bits.

$$M \cdot N \cdot b = C. \tag{1}$$

Determine M, N, b such that the following error is minimum.

$$E = \iint_{\Omega} [f(x,y) - \tilde{f}(x,y)]^2 \, dx \, dy \Big/ \iint_{\Omega} dx \, dy. \tag{2}$$

A number of restrictions are introduced to make the problem tractable analytically. The function f is characterized by its value range $R = \max f - \min f$, and the mean fluctuation rates σ_x and σ_y defined by

$$\sigma_x^2 = \overline{(f_x')^2} \quad \text{and} \quad \sigma_y^2 = \overline{(f_y')^2}. \tag{3}$$

The quantization error n is defined as $n = Q\hat{f} - \hat{f}$. Zero-order hold interpolation is used. This means that $\tilde{f}(x, y) = Q\hat{f}(x_i, y_j)$ for x, y around x_i, y_j.

III. Solution

The criterion (2) is expanded by dividing the image in $M \times N$ cells with sides $\delta_x = L_x/M$ and $\delta_y = L_y/N$. The cell midpoint x_i, y_j belongs to the grid G. The contributions from all cells are then summed up. Insertion of the digitization and the interpolation schemes in the criterion (2) gives the mean square error

$$\bar{E} = \frac{1}{L_x L_y} \sum_{i,j} \left\{ \iint_{\square} \overline{[f(x - x_i, y - y_j) - f(x_i, y_j)]^2} \, dx \, dy + \iint_{\square} \overline{n(x_i, y_j)^2} \, dx \, dy \right\} \tag{4}$$

where $\square$ denotes integration over one cell. Steiglitz calls the first term in (4) the reconstruction error. This error depends only on the grid resolutions δ_x and δ_y. A Taylor series of f gives the following expression for the reconstruction error in one cell.

$$\frac{1}{12} \cdot (\delta_x^3 \delta_y \sigma_x^2 + \delta_x \delta_y^3 \sigma_y^2). \tag{5}$$

The second term in (4) is the quantization error, which depends only on the number of quantization levels 2^b. Assuming equidistant quantization with the grain $\delta = R \cdot 2^{-b}$, the quantization error is approximated by $\delta^2/12$ and

$$\iint_{\square} \overline{n(x_i, y_j)^2} \, dx \, dy = \delta_x \delta_y \overline{n^2} = \delta_x \cdot \delta_y \cdot \frac{1}{12} \cdot \frac{R^2}{2^{2b}}. \tag{6}$$

The following formula is then obtained for the total mean square error (4).

$$\bar{E} = \frac{1}{12} \cdot \left(\frac{L_x^2 \sigma_x^2}{M^2} + \frac{L_y^2 \sigma_y^2}{N^2} + \frac{R^2}{2^{2b}} \right). \tag{7}$$

The optimal digitization problem is to minimize (7) subject to the constraint (1). The variables L_x, σ_x, L_y, σ_y, and R are known constants which depend on the image. The problem is solved simply by completing the squares in (7) and inserting the constraint (1). The solution is

$$b = \frac{1}{2 \ln 2} \ln \left[C \cdot \ln 2 \cdot \frac{R^2}{L_x \sigma_x L_y \sigma_y} \right] \tag{8}$$

$$M = \sqrt{\frac{L_x \sigma_x}{L_y \sigma_y}} \cdot \sqrt{\frac{C}{b}} \tag{9}$$

$$N = \sqrt{\frac{L_y \sigma_y}{L_x \sigma_x}} \cdot \sqrt{\frac{C}{b}}. \tag{10}$$

It is interesting to see how the relation between the value range R and the fluctuation rates σ_x and σ_y influence the solution. More fluctuations leads to fewer bits (lower b) and more image resolution. Fewer fluctuations requires more bits and fewer samples. This agrees with earlier subjective tests [3], [4].

IV. The Experiment

The criterion (2) was chosen largely for mathematical convenience. A number of experiments have been carried out to see if the optimum digitization obtained corresponds to the subjective notation of a good digitization [7]. Fig. 1 illustrates with an original image and a number of digitized versions. A scan of the original image gives the characteristics

$$R = 4.06 \cdot 10^{-2}\, L_x \sigma_x = 9.68 \cdot 10^{-2}\, L_y \sigma_y. \tag{11}$$

It is seen from expression (11) that there are more fluctuations horizontally than vertically. The number of bits used for digitization is $C = [256]^2 = 65\,536$. The product $C = M \cdot N \cdot b$ cannot be kept constant, since M, N, and b are integers. The hardware also limits the possible values on M and N to 512, 256, 170, 128, 102, 85, $\cdots$ (= trunc. $(512/k)$). The integers which are closest to the optimal are $M = 170$, $N = 85$, and $b = 4$ [Fig. 1(c)]. If fewer bits/pixel ($b < 4$) are used, the image looks blurred [Fig. 1(b)]. If more bits/pixel ($b > 4$) and fewer sampling points are used, some detail is lost. If the optimal number of bits ($b = 4$) are used but less care is taken to the two directions ($M = 128$ and $N = 102$), it gives an "edgier" image because some detail is lost horizontally but not much is gained vertically [7].

V. Conclusions

A theoretical formulation of the optimal digitization problem is given. The solution is obtained from an optimization of a criterion due to the constraint of fixed number of bits. The solution is tested experimentally and agrees well with human visual quality. An advantage is that the solution is given in closed form [see (8)–(10)]. This makes it easy to use as a rule of thumb. It also clearly points out the dependence on image characteristics. This dependence explains and agrees with results of other subjective tests. The criterion (2) is one of the simplest which admits an analytic solution. It would be interesting to look at other alternatives and make more extensive experimentation.

Acknowledgment

We are grateful to Dr. T. Pavlidis who suggested the problem to the third author. Support from the Swedish Board of Technical Development and the National Science Foundation is gratefully acknowledged.

References

[1] W. K. Pratt, *Digital Image Processing.* New York: Wiley, 1978.
[2] T. Pavlidis, *Algorithms for Graphics and Image Processing.* Rockville, MD: Comput. Sci. Press, 1982, p. 39.
[3] A. Rosenfeld and A. C. Kak, *Digital Picture Processing.* New York: Academic, 1982, p. 111.
[4] T. S. Huang, O. J. Tretiak, B. Prasada, and Y. Yamaguchi, "Design considerations in PCM transmission of low resolution monochrome still pictures," *Proc. IEEE*, vol. 55, pp. 331–335, 1967.
[5] I. E. Abdou and K. Y. Wong, "Analysis of linear interpolation schemes for bi-level image applications," *IBM J. Res. Develop.*, vol. 26, no. 6, 1982.
[6] K. Steiglitz, "Transmission of an analog signal over a fixed bit-rate channel," *IEEE Trans. Inform. Theory*, vol. IT-12, no. 4, 1966.
[7] L. Nielsen, K. J. Aström, and E. I. Jury, "Optimal digitization of 2-D images," CODEN:LUTFD2/(TFRT-7265)/1-023/(1983), 1983.

In 1984:

L. Nielsen, KJ Aström & E.I. Jury
published a paper titled:

OPTIMAL DIGITIZATION OF 2-D IMAGES

(IEEE T-ASSP-32/6, 1984)

The list of authors is very interesting

- K. J. Aström - a world famous specialist in ADAPTIVE CONTROL THEORY
  Lund, Sweden
- L. Nielsen - professor of vehicular systems
  (student - first paper in 1984) Linköping U. Sweden
  (visual servoing for vehicles!)
- E. I. Jury - world famous in Control Theory
  Z-transform (Stability Criteria etc)
  U. of Miami, Coral Gables
  FLORIDA, U.S.A.

The paper has an ACKNOWLEDGMENT:

WE ARE GRATEFUL TO

Dr Theo PAVLÍDIS

WHO SUGGESTED THE PROBLEM TO THE THIRD AUTHOR (E. I. Jury!)

Theo Pavlidis, a leader in Image Processing & Analysis & Graphics was at that time involved with

SYMBOL TECHNOLOGIES

and was interested in a wealth of "practical problems".

The Nielsen Astrom Jury paper was subsequently quoted about 5/6 times, of these quotations by AMB + coauthors about 3/4 times.

- AMB * On Optimal Image Digitization
  IEEE T ASSP, 1987 (Vol 35/4)

- N Kiryati & AMB & A. Jonas
  Bit Allocation in Piecewise Planar Representation of Images
  J. Vis Comm & Imag Representation 6/1 1995

---------- *

- N. Kiryati & AMB
  Gray Levels Can Improve the Performance of Binary Image Digitizers
  CVGIP: GMP 53, 1991

- AMB, M Elad & R Kimmel
  Down Scaling for Better Transform Compression
  IEEE Trans I.P. 12/9, 2003

* While writing these papers AMB corresponded with T. Pavlidis.

The Problem:

Given a set of signals $f_\omega(x)$ over $[0,1]$ with values over $[-1,1]$, represent/describe them as best you can with less than $\underline{B}$ bits.

Solution

• Sample then • Quantize

(How to sample? How to quantize?)

If $\underline{N}$ samples will be quantized to $\underline{b}$ bits each we'll have to have

$$\underline{N} \cdot \underline{b} \leq \underline{B}$$

How good is a representation

is measured by a distance between

$f_\omega(x)$ and $f_\omega^{S(N)+Q(b)}(x)$ ← estimated from samples quantized.

THE PROTOTYPE PROBLEM

We want to solve is:

$$\boxed{\min_{N,b} \text{Distance}\left[f_{\omega}^{S(\underline{N})+Q(b)}(x) \, , \, f_{\omega}(x) \right] \quad \text{subject to } N \cdot b \leq B}$$

$\text{Distance}\left[f_{\omega}^{S(N)+Q(b)}(x) \, , \, f_{\omega}(x) \right]$ will be an expression involving $\underline{N}$, $\underline{b}$ and the properties of the family of functions $\{f_{\omega}(x)\}$

LET US TAKE THE 1-D EXAMPLE OF Nielsen Astrom A Jury, and LOOK AT IT.

GIVEN A FUNCTION $f(x)$ over $[0,1]$

• SAMPLE BY CONSIDERING N EQUAL SIZED INTERVALS OF LENGTH $\delta = 1/N$, $R_i = [i/N, i+1/N)$ $i = 0, 1, \ldots N-1$ and DESCRIBE $f(x)$ over R_i by a SINGLE NUMBER $f(i)$.

• QUANTIZE THE NUMBERS $f(i)$ by selecting 2^b values in the range $[-1, 1]$ and mapping $f(i)$ to $f^Q(i)$ "the closest" of the 2^b values.

Then

$$\boxed{f^{S(N)+Q(b)}(x) \triangleq f^Q(i) \text{ for } x \in R_i}$$

s "a piecewise constant representation of $f(x)$ by N.b bits."

Let us look at the "mean square" distance between $f(x)$ and its representation

$$D\left(f(x), f^{S(N)+Q(b)}_{(x)}\right) \triangleq \int_0^1 \left[f(x) - f^{S+Q}_{(x)}\right]^2 dx$$

$$= \sum_{i=0}^{N-1} \int_{R_i} \left[f(x) \underbrace{- f(i) + f(i)} - f^{Q}_{(i)}\right]^2 dx =$$

$$= \sum_{i=0}^{N-1} \int_{R_i} \left[f(x) - f(i)\right]^2 dx +$$

$$\sum_{i=0}^{N-1} 2\left(f(i) - f^{Q}_{(i)}\right) \int_{R_i} \left(f(x) - f(i)\right) dx + \quad 0$$

here we see that selecting
$f(i)\int_{R_i} dx \equiv \int_{R_i} f(x)dx$
we get 0 cross terms

$$+ \sum_{i=0}^{N-1} \left(f(i) - f^{Q}_{(i)}\right)^2 \int_{R_i} dx$$

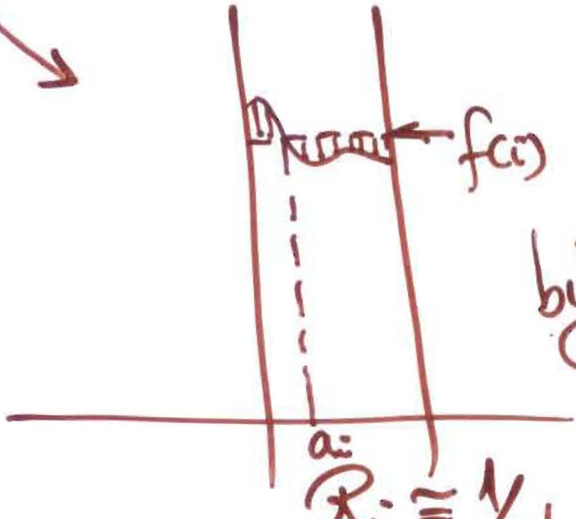
Therefore we get

$$D\left(f(x), f^{S+Q}_{(x)}\right) =$$

$$\underbrace{\sum_{i=0}^{N-1} \int_{R_i} \left[f(x) - \frac{1}{\delta}\int_{R_i} f(\xi)\,d\xi\right]^2}_{\text{"SAMPLING RECONSTRUCTION ERROR"}} + \underbrace{\sum_{i=0}^{N-1} \delta \left(f(i) - f^{Q}_{(i)}\right)^2}_{\text{AVERAGE QUANTIZATION ERROR}}$$

$$\frac{1}{N}\sum_{i=0}^{N-1}\left(f(i) - f^{Q}_{(i)}\right)^2$$

$$\cong \boxed{1/2^{2b}}$$



$R_i \cong 1/N$

by Taylor Expansion here

$$\cong \frac{1}{N^2}\left(\underbrace{\frac{1}{N}\sum_i \dot{f}^2(a_i)}_{\sigma_f}\right)$$

Hence we obtained that

$$D\left(f(x), f^{S+Q}(x)\right) \cong \frac{\sigma_f}{N^2} + \frac{K^Q}{2^{2b}}$$

$\left\{ \text{if we quantize optimally and select best representations over } R_i \text{ to be quantized.} \right\}$

Now we can do:

$$\begin{cases} \min & \dfrac{\sigma_f}{N^2} + \dfrac{K^Q}{2^{2b}} \\ \text{s.t.} & N \cdot b = B \end{cases}$$

CUTE ISN'T IT!

So DO:

$$\underset{(N,b)}{\Psi} \triangleq \left(\frac{\sigma_f}{N^2} + \frac{K^Q}{2^{2b}}\right) + \lambda\left(B - N b\right)$$

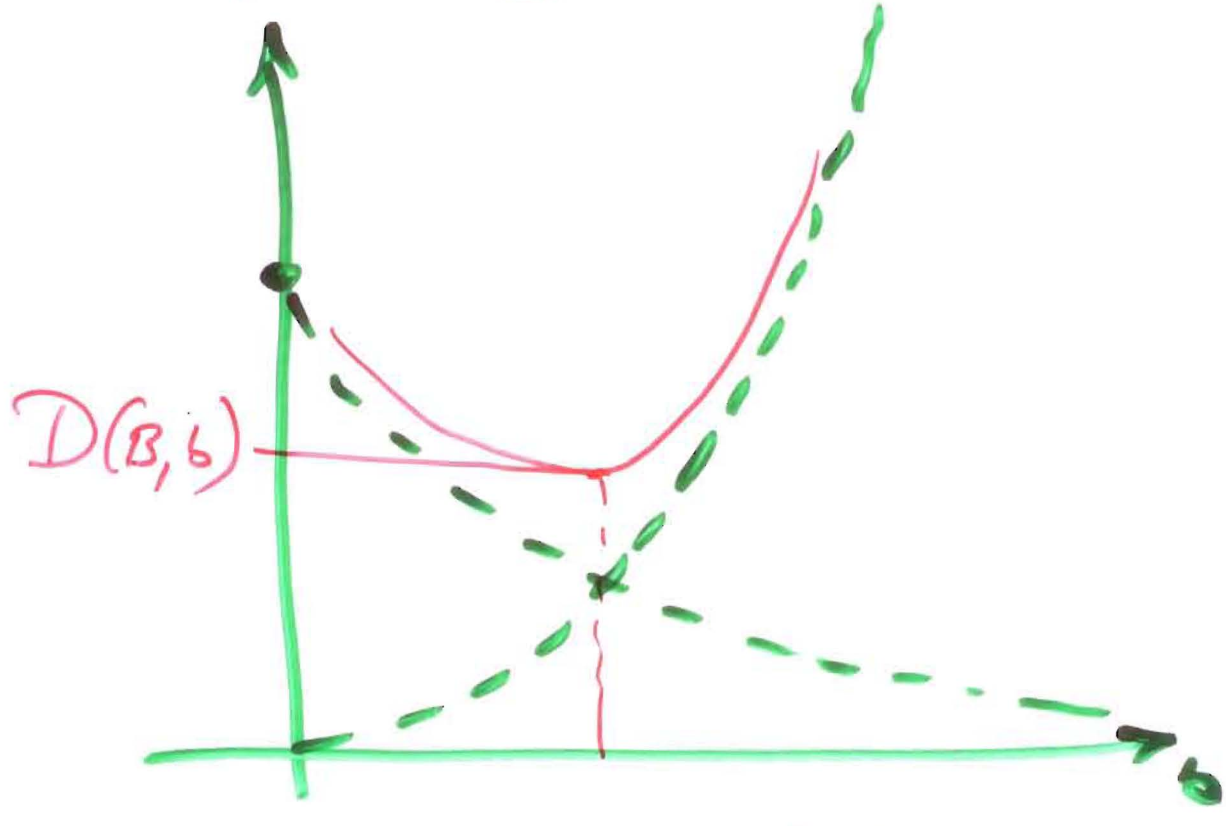
$\frac{\partial \Psi}{\partial N} = 0 \quad \frac{\partial \Psi}{\partial b} = 0$ and get

SOLUTIONS for N, b

$$D(B,b) = \frac{\sigma_f}{\left(\frac{B}{b}\right)^2} + \frac{K^Q}{2^{2b}} =$$

$$= \underbrace{\frac{\sigma_f}{B^2} \cdot b^2}_{\text{grows}} + \underbrace{\frac{K^Q}{2^{2b}}}_{\text{decreases}}$$

as $b \nearrow$



etcetera

etcetera

etcetera

...

This was just an example, somewhat similar to what NAJ did. There are many variations possible.

Obviously: "SAMPLING" can be done in many ways. Define a family of functions $\{\varphi_i\}$ and project $f(x)$ on these via

$$\langle f(x), \varphi_i \rangle = f_i$$

Then f_i are "generalized" samples of f. (Now φ_i can be Wavelets or even dictionaries!).

If $\{\varphi_i\}$ are O.N. then we have

$$f^{S+Q}(x) = \sum_{i=1}^{N} f_i^{Q} \varphi_i(x) \text{ and the error is}$$

$$D\left[f(x), f^{S+Q}(x)\right] = \int_{[0,1]} \left[f(x) - f^{S+Q}(x)\right]^2 dx =$$

$$= \int_{[0,1]} \left[f(x) - \sum_{i=1}^{N} f_i^{Q} \varphi_i(x)\right]^2 dx =$$

$$= \int_{[0,1]} \left(f(x) - \sum^{N} f_i \varphi_i(x) + \sum^{N} (f_i - f_i^{Q}) \varphi_i(x)\right)^2 dx$$

$$= \underbrace{\int_{[0,1]} \left[f(x) - \sum^{N} f_i \varphi_i(x)\right]^2 dx}_{\text{Representation Error}} +$$

$$+ 2 \int \left[f(x) - \sum^{N} f_i \varphi_i(x)\right] \left[\sum^{N} (f_i - f_i^{Q}) \varphi_i(x)\right] dx$$

$$\sum^{N} (f_i - f_i^{Q}) \cdot \underbrace{\int (f(x)\varphi_i(x) - f_i)}_{0}$$

$$+ \underbrace{\sum^{N} (f_i - f_i^{Q})^2}_{\text{Quantization Error.}} \underbrace{\int \varphi_i^2(x) dx}_{1}$$

- Our work on JPEG!
  - Our work on piecewise planar representation.
  - Our work on encoding B/W images.

all these can be regarded as processes of SAMPLING and QUANTIZATION.

Question:

Where did Nyquist dissappear in all this!